# ENUMERATION OF TREES

*Peter Kirschenhofer*
Inst. f. Algebra u. Diskrete Mathematik
Technische Universität Wien
A - 1040 Vienna, Austria

*Table of Contents:*

## 1. INTRODUCTION

Enumeration of trees is a more and more rapidly growing area of enumerative Combinatorics, especially since a number of problems in Computer Science, e.g. in the average case analysis of data structures and algorithms, involve the task to enumerate trees of a specified kind. It is the aim of this article to survey some of the most important methods and results in this area. Of course by no means we can give a complete overview. Nevertheless we hope that the selection of problems and methods might be helpful.

This article contains explicit combinatorial enumeration formulae as well as asymptotic results: in fact it turns out that in many problems of practical interest the latter kind of results is either the only one that can be achieved or even the preferable one for the interpretation of final results.

The author would like express his deep gratitude to *Ph. Flajolet* (INRIA, Rocquencourt) for his pioneering contributions to the application of enumerative and asymptotic methods in the analysis of algorithms and numerous stimulating discussions on the author's work within this subject.

## 2. SOME BASIC RESULTS IN UNLABELLED TREE ENUMERATION

### 2.1 Preliminaries on the Combinatorics of the Ordinary Generating Function (o.g.f.)

In order to give precise definitions for the families of trees we will be concerned with it is very helpful to use the *Flajolet* notion of the operator method for o.g.f. ([17]):

Let $\mathcal{A},\mathcal{B},\ldots$ be families of combinatorial objects with weight functions $|\cdot|_{\mathcal{A}}$, $|\cdot|_{\mathcal{B}}$, ... where the weights are natural numbers. (The reader might think e.g. of $\mathcal{A}$ as the family of all trees and of $|t|_{\mathcal{A}}$ as the number of nodes of tree t.) By $\mathcal{A}_n$, $\mathcal{B}_n$,... we denote the objects in $\mathcal{A},\mathcal{B},\ldots$, with weight equal to n. We assume that each of these families is finite and set

$$a_n = |\mathcal{A}_n|,\ b_n = |\mathcal{B}_n|,\ldots .$$

The o.g.f. of $\mathcal{A}$ is

$$A(z) = \sum_n a_n z^n = \sum_{t\in \mathcal{A}} z^{|t|_{\mathcal{A}}}. \tag{2.1}$$

The basic idea of the operator method is to associate with a certain combinatorial construction

$$\Phi(\mathcal{A},\mathcal{B},\ldots)$$

in the area of objects an operator

$$\Psi(A(z),B(z),\ldots)$$

in the area of o.g.f.

The most important combinatorial constructions which translate into an operator on o.g.f. are summarized in the following:

1) $C = A \dot\cup B$ (*disjoint union*) with $|t|_C = \begin{cases} |t|_A & \text{if } t \in A \\ |t|_B & \text{if } t \in B \end{cases}$

corresponds to the *sum* of o.g.f.:

$$C(z) = A(z)+B(z). \tag{2.2}$$

2) $C = A \times B$ (*Cartesian product*) with $|t|_C = |(t_1,t_2)|_C = |t_1|_A + |t_2|_B$

corresponds to the *Cauchy product* of o.g.f.:

$$C(z) = A(z)\cdot B(z) \tag{2.3}$$

2') $C = A^k$ corresponds to $C(z) = A(z)^k$.

3) $C = A^*$ (finite *sequences* of objects from A), where $a_0=0$ and $A^* = \dot\bigcup_{k\geq 0} A^k$, with

$A^\circ = \{\varepsilon\}$, $|\varepsilon|=0$, corresponds to the *geometric series*

$$C(z) = \frac{1}{1-A(z)} \ . \tag{2.4}$$

4) $C = M[A]$ (*multisets* of objects from A), where $a_0=0$, corresponds to

$$C(z) = \exp(A(z) + \frac{A(z^2)}{2} + \frac{A(z^3)}{3} + \ldots). \tag{2.5}$$

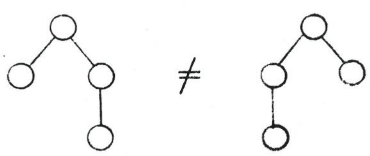
## 2.2 Planted Plane Trees

The family $P$ of *planted plane trees* or ordered trees is defined by

$$\begin{aligned} P &= \{o\} \times (\{\varepsilon\} \dot\cup P \dot\cup P^2 \dot\cup \ldots) \\ &= \{o\} \times P^* \end{aligned} \tag{2.6}$$

i.e. a planted plane tree consists of a root followed by (eventually zero) subtrees, where the relative order of the subtrees is relevant.

Example:



The weight $|t|_P$ is the number of nodes of t.

From Section 2.1 we know that equation (2.6) translates into

$$P(z) = \frac{z}{1-P(z)} \ , \text{ where } P(0)=0, \tag{2.7}$$

so that

$$P(z) = \frac{1-\sqrt{1-4z}}{2} \ . \tag{2.8}$$

The number $p_n$ of n-node planted plane trees may now be computed either from (2.7) using the Lagrange Inversion formula (compare e.g. [22]) or from (2.8) using the binomial series:

$$p_n = \frac{1}{n}\binom{2n-1}{n-1}, \tag{2.9}$$

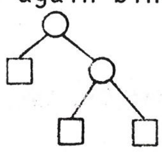
a *Catalan number*.

## 2.3 (Extended) Binary Trees

The family $\mathcal{B}$ of (extended) *binary trees* is defined by

$$\mathcal{B} = \{\square\} \mathbin{\dot\cup} \{o\} \times \mathcal{B}^2, \tag{2.10}$$

i.e. a tree in $\mathcal{B}$ is either a single leaf $\square$ or a root o followed by a left and a right subtree, which are again binary trees, e.g.



By $|t|_{\mathcal{B}}$ we denote the number of internal nodes o in t. From (2.10) we get for the o.g.f.:

$$B(z) = 1 + z\cdot B(z)^2, \tag{2.11}$$

so that

$$B(z) = \frac{1-\sqrt{1-4z}}{2z} \tag{2.12}$$

or

$$b_n = \frac{1}{n+1}\binom{2n}{n} = p_{n+1}, \tag{2.13}$$

i.e. the number of binary trees with n internal nodes equals the number of planted plane trees with n+1 nodes in total. We will present a bijective proof for this fact below.

For the moment let us mention that from (2.10) we also get the double o.g.f. $B(z,u)$ where z marks internal nodes and u marks leaves:

$$B(z,u) = u + z\cdot B(z,u)^2 \tag{2.14}$$

so that

$$B(z,u) = \frac{1-\sqrt{1-4zu}}{2z} = u\cdot B(zu). \tag{2.15}$$

It follows that the numbers $b(n,m)$ of binary trees with n internal nodes and m leaves are given by

$$b(n,m) = b_n\cdot\delta_{m,n+1}, \tag{2.16}$$

i.e. *a binary tree with n internal nodes has n+1 endnodes.*

There is no such simple correspondence for general planted plane trees: From equation (2.6) we find

$$P(z,u) = z(u+P(z,u)+P(z,u)^2+\dots)$$
$$= z(u-1+(1-P(z,u))^{-1}), \qquad (2.17)$$

or

$$P(z,u) = (1-z(1-u))\cdot P\left(\frac{zu}{(1-z(1-u))^2}\right) , \qquad (2.18)$$

but it is not that easy to derive an explicit formula for the quantities $p(n,m)$ from (2.18). We will return to that question in Section 2.5.

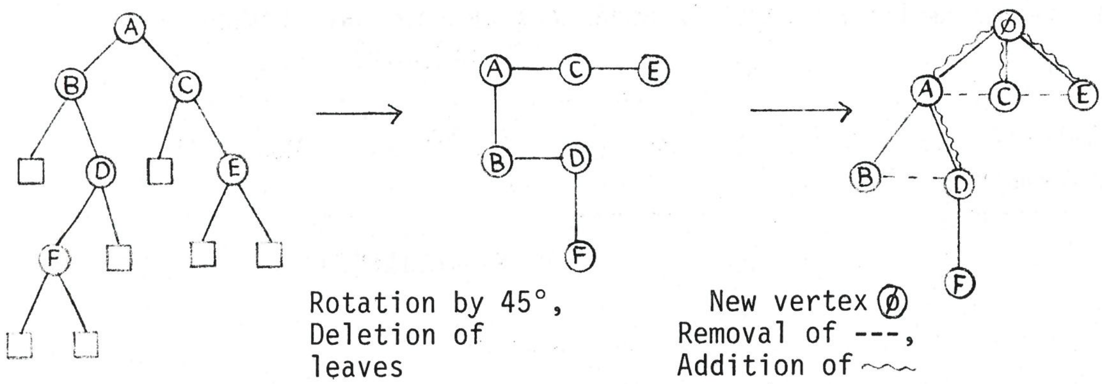
## 2.4 The Rotation Correspondence

In 2.3 we have mentioned that

$$b_n = p_{n+1} \quad . \qquad (2.19)$$

A bijective proof for this fact can be given using the *"Rotation Correspondence"*: Starting from a tree in $\mathcal{B}_n$, i.e. a binary tree with n internal nodes, in a first step we rotate the tree by 45 degrees and delete the leaves. In the second step we introduce a new node $\emptyset$ which shall become the root of the outcoming planted plane tree. Then we remove all horizontal edges and add new vertical edges between i) $\emptyset$ and all nodes in the upper level and ii) all nodes that were connected by horizontal edges and their common "ancestor" in the next higher level.

Example:




Rotation by 45°, Deletion of leaves

New vertex $\emptyset$, Removal of ---, Addition of ~~~ edges

It is easily seen that these operations create a planted plane tree with n+1 nodes in total and that the mapping defines a bijection between $\mathcal{B}_n$ and $\mathcal{P}_{n+1}$.

Let the *level* of a node of a planted plane tree be its distance from the root, and the *left-sided level* of a node of a binary tree its "left-sided distance" from the root, i.e. the number of all left-directed edges on the unique path between the root and this node.

The Rotation Correspondence transforms an internal node of left-sided level $l+1$ into a node of level $l$. In other words: *The left sided "height" of a tree* $t \in \mathcal{B}_n$ *equals the height of the corresponding planted plane tree* $\in P_{n+1}$. (As usual we denote by the height the maximum level of a node).

There is also an *important algorithmic interpretation* for the Rotation Correspondence, compare *Knuth* [37]: A binary tree may be *"traversed" recursively* using one of the following three fundamental principles:

a) *Preorder traversal*: Visit the root (I)
Traverse left subtree (II) (2.20)
Traverse right subtree (III)

b) *Inorder traversal*: II, I, III (2.21)

c) *Postorder traversal*: II, III, I. (2.22)

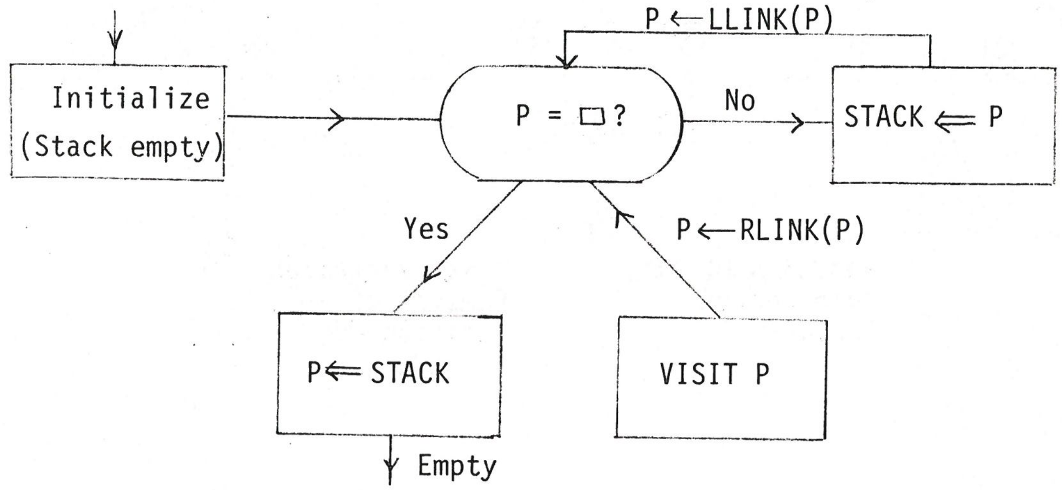
<u>Example:</u> Considering the binary tree from the last example above the internal nodes will be "visited" in the following order:

Preorder: ABDFCE

Inorder: BFDACE

Postorder: FDBECA

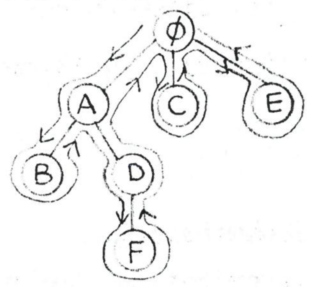
Knuth [37] presents the following *algorithm for Inorder traversal*, where P is a pointer and an auxiliary stack is used to keep necessary nodes:



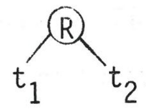

<u>Example:</u> The binary tree t from the last example produces the stack sequence

(ø), A, AB, A, AD, ADF, AD, A, ø, C, ø, E, ø
B F D A C E

This sequence is also produced by following the contours of the planted plane tree associated with t via the rotation correspondence:

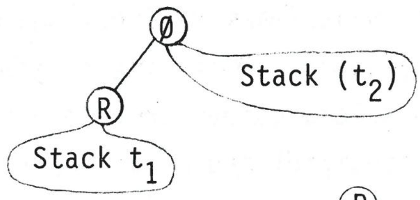



In general the stack sequence of the tree

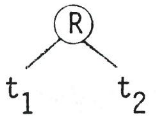



is $\left\{\begin{array}{l} R \\ R\ \text{Stack}\ (t_1) \\ \emptyset \longrightarrow (R) \\ \text{Stack}\ (t_2) \end{array}\right.$

Thus the stack sequence is gained by following the contours of the planted plane tree

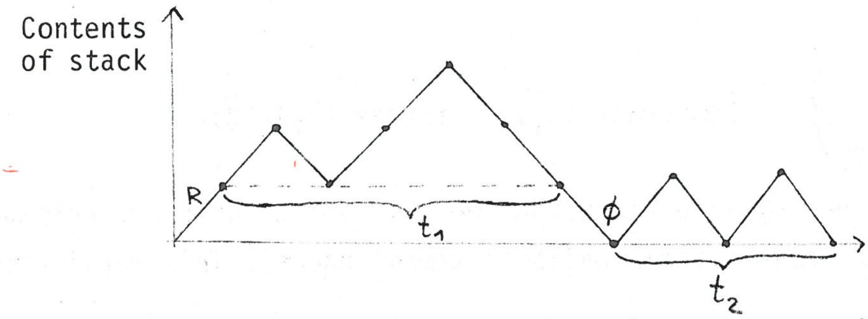



which is the tree corresponding to




via the Rotation Correspondence.

If we depict the contents of the stack as a random walk, we get in the above example:




This is the well-known correspondence between trees in $P_{n+1}$ and non-negative lattice paths starting with (0,0) and ending with (2n,0).

We mention in passing that the last bijection allows to find the explicit formula (2.9) using the reflection principle of *D. André* (compare e.g. [8]).

## 2.5 The Cycle Lemma

The *Cycle Lemma* of *Dvoretzky* and *Motzkin* [14] is a well-suited instrument for several tree enumeration problems. A recent paper on this subject is due to *Dershowitz* and *Zaks* [12].

A sequence $p_1p_2...p_l$ of boxes and balls is called *k-dominating* if for every position i, $1 \leq i \leq l$, the number of boxes in $p_1p_2...p_i$ is more than k-times the number of balls.

Example:

□ □ □ □ □ ○○ □ □ ○ is 2-dominating,

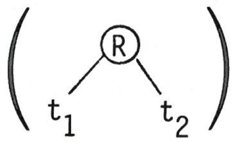
□ □ □ □ ○○ □ □ □ ○ is 1-dominating, but not 2-dominating.

Cycle Lemma: For any sequence $p_1p_2...p_{m+n}$ of m boxes and n balls, where $m \geq kn$, there exist exactly m-kn cyclic permutations that are k-dominating.

Sketch of proof: Arrange $p_1...p_{m+n}$ on a circle. The removal of k boxes followed by 1 ball, i.e. of the pattern $\square^k0$, does not change the number of k-dominating permutations.

As long as the number of boxes is at least k-times the number of balls and the latter >0 there must be a subsequence $\square^k0$ by the pigeon-hole principle. Successive removal of subsequences $\square^k0$ lets us end up with a sequence of m-kn boxes which correspond to the (beginnings) of the m-kn cyclic permutations, that yield k-dominating sequences.

As *a first application of the Cycle Lemma* we give *another proof for* $|B_n| = b_n$: Traverse the trees in $B_n$ in postorder (2.22), and note the leaves (□) and internal nodes (o) in the order they are visited. By the definition of postorder we have

Postorder  = Postorder ($t_1$) Postorder ($t_2$) Ⓡ ,

so that it is easily seen by induction, that we will end up with a 1-dominating sequence of n+1 boxes (leaves) and n balls (internal nodes). This map is bijective:

Let $p_1p_2...p_kp_{k+1}...p_{2n+1}$ be a 1-dominating sequence of n+1 boxes and n balls. In order to ensure that we can uniquely reconstruct a tree in $B_n$ from this sequence (starting from the end which must denote the root,...) we only have to prove that for no k ($1 \leq k \leq 2n$) the number of boxes in $p_{k+1}...p_{2n+1}$ may exceed the number of balls. But if this should happen, we had the implication that the number of boxes in $p_1...p_k$ is less or equal to the number of balls, a contradiction to the fact that $p_1...p_{2n+1}$ is 1-dominating.

Therefore we have shown that $|B_n|$ equals the number of 1-dominating sequences of n+1 boxes and n balls. By the Cycle Lemma exactly (n+1)-n=1 of the 2n+1 cyclic permutations of any of the $\binom{2n+1}{n+1}$ sequences of n+1 boxes and n balls is 1-dominating. Thus

$$|\mathcal{B}_n| = \frac{1}{2n+1}\binom{2n+1}{n+1} = \frac{1}{n+1}\binom{2n}{n},$$

as in (2.13).

A similar argument holds for *t-ary trees*:

The family $\mathcal{T}$ is defined by

$$\mathcal{T} = \{\square\} \mathbin{\dot\cup} \{o\} \times \mathcal{T}^t, \tag{2.23}$$

i.e. a t-ary tree has each internal node followed by exactly t subtrees. The weight of a tree in $\mathcal{T}$ is the number of internal nodes. The o.g.f. fulfills

$$T(z) = 1 + z \cdot T^t(z) \; ; \tag{2.24}$$

if we mark leaves by u we get

$$T(z,u) = u + z \cdot T^t(z,u), \tag{2.25}$$

so that

$$T(z,u) = u \cdot T(zu^{t-1}). \tag{2.26}$$

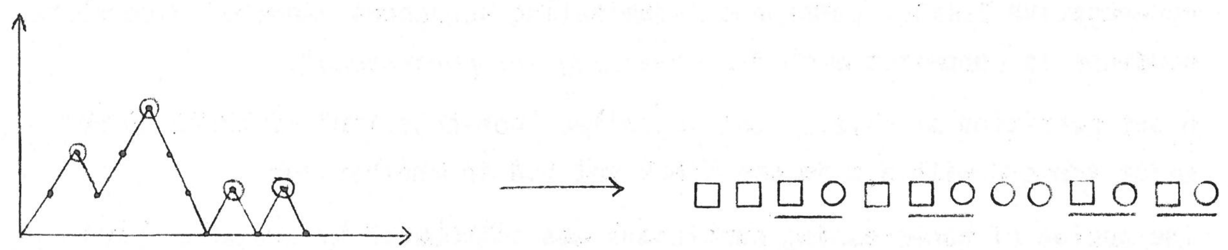
Thus a tree in $\mathcal{T}_n$ has (t-1)n+1 leaves.

Postorder traversal of a tree in $\mathcal{T}_n$ yields a (t-1)-dominating sequence of (t-1)·n+1 boxes (leaves) and n balls (internal nodes). By the Cycle Lemma there is ((t-1)·n+1)-(t-1)·n=1 (t-1)-dominating cyclic permutation of any sequence of (t-1)n+1 boxes and n balls. Therefore we find

$$|\mathcal{T}_n| = \frac{1}{tn+1}\binom{tn+1}{n} = \frac{1}{(t-1)n+1}\binom{tn}{n}. \tag{2.27}$$

We have mentioned in Section 2.3 that the number of planted plane trees in $\mathcal{P}_n$ with m leaves cannot be seen immediately from the corresponding o.g.f.(2.18). Nevertheless the Cycle Lemma allows to compute these numbers as well:

Let us consider the family $\mathcal{P}_{n+1,m}$ of trees in $\mathcal{P}_{n+1}$ with m leaves. Following the contours of a tree in $\mathcal{P}_{n+1,m}$ we get a nonnegative random walk connecting (0,0) and (2n,0) (compare 2.4). Now we assign a box to each upward step and a ball to each downward step and add one additional box at the beginning:



We get a 1-dominating sequence of n+1 boxes, n balls and m subblocks □o (since each of these subblocks corresponds to a peak of the random walk, i.e. to a leaf of the tree).

The total number of sequences of n+1 boxes, n balls and m □o's starting with □ and ending with o is $\binom{n}{m-1} \cdot \binom{n-1}{m-1}$.

By the Cycle Lemma there is (n+1)-n=1 cyclic permutation of each such sequence which is 1-dominating, whereas in total there are m cyclic permutations that transform the sequence into another one of the same type.

Altogether we get:

$$p_{n+1,m} = |P_{n+1,m}| = \frac{1}{m}\binom{n}{m-1}\binom{n-1}{m-1} = \frac{1}{n}\binom{n}{m}\binom{n}{m-1}, \tag{2.28}$$

the *NARAYANA-numbers* (compare [45]; the proof using the Cycle Lemma may be found in [10]).

An immediate consequence of (2.28) is:

$$p(n+1,m) = p(n+1,n+1-m), \tag{2.29}$$

i.e. *the number of trees in $P_{n+1}$ with m leaves equals the number of trees in $P_{n+1}$ with m internal nodes!*

A *bijective proof* for the last observation may be given as follows: Starting from a tree in $P_{n+1,m}$ we apply the Rotation Correspondence and get a tree in $B_n$ with m "left-sided" leaves. "Reflecting" this binary tree, i.e. interchanging all left and right edges, we get a tree in $B_n$ with n+1-m left-sided leaves (since there are n+1 leaves in total). The inverse Rotation Correspondence finally transforms this binary tree into a tree in $P_{n+1,n+1-m}$.

Another consequence of above is the following:
The *average number of leaves of a tree in $P_{n+1}$* equals the average number of left-sided leaves of a tree in $B_n$, which is (by reflection) obviously $\frac{n+1}{2}$.

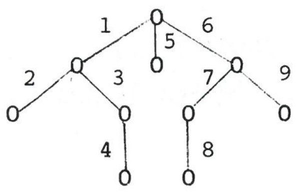
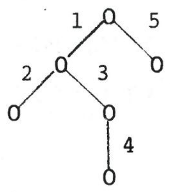
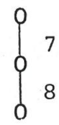
## 2.6 Planted Plane Trees and non-crossing Partitions

So far we have noted bijections between planted plane trees and binary trees, non-negative lattice paths and 1-dominating sequences. Another nice correspondence is concerned with *"non-crossing set partitions"*.

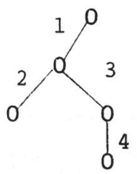
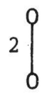
A set partition of $\{1,2,\dots,n\}$ is called "non-crossing" if there do not exist a<b<c<d with a,c in one block and b,d in another one.

The notion of non-crossing partitions was introduced by *Kreweras* [40] and further investigated by *Poupard,* and *Edelman* [15]. *Prodinger* [48] gave the following bijection between non-crossing partitions of $\{1,2,...,n\}$ into m blocks and $P_{n+1,m}$ (Another bijection is given in [11]):

Starting from a tree in $P_{n+1,m}$ we label the edges according to their first occurrence in preorder traversal of the tree. Then we remove the path connecting the root with the edge labelled "n"; the numbers of this path create the first block of the partition:


→

,

, $\{6,9\}$

This procedure is recursively repeated with the remaining trees, and, by the construction, we end up with a non-crossing partition

→

, $\{5\},\{7,8\},\{6,9\}$ →

, $\{1,3,4\},\{5\},\{7,8\},\{6,9\}$

→ $\{2\},\{1,3,4\},\{5\},\{7,8\},\{6,9\}$

The inverse mapping starts with the block containing "n" and forms a chain with its elements in monotone order. Then, recursively, we take the block with highest number not yet used and attach its first edge "a" from the left

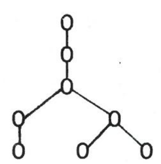
- between "b" and "c" with b<a<c if such b and c exist in the tree so far
- at the root, otherwise.

## 2.7 Motzkin Trees

A *Motzkin tree* (or *unary-binary* tree) is a tree in family $M$ defined by

$$M = \{0\} \times (\{\varepsilon\} \cup M \cup M^2) \tag{2.30}$$

Example:



The weight of a tree in $M$ is the number of its nodes. From (2.30) we get for the o.g.f.

$$M(z) = z\cdot(1+M(z)+M^2(z)), \tag{2.31}$$

so that

$$M(z) = \frac{1-z-\sqrt{1-2z-3z^2}}{2z}, \tag{2.32}$$

and

$$m_n = |M_n| = \sum_{j \geq 0} \frac{1}{j+1} \binom{2j}{j} \binom{n-1}{2j}, \tag{2.33}$$

the *Motzkin numbers*.

It is easily seen that $M_{n+1}$ corresponds to the nonnegative random walks connecting (0,0) and (n,0) where a step may lead one unit upwards or downwards or at the same level to the right.

Another construction of $M$ is from the binary trees $B_0$ in $P$ (i.e. binary trees where internal nodes and leaves are not distinguished in the weight):
If we substitute the nodes of the trees in $B_0$ by chains of finite length $\geq 1$, we create all Motzkin trees:

This fact translates into the following equation for the o.g.f.:

$$M(z) = B_0\left(\frac{z}{1-z}\right). \tag{2.34}$$

## 2.8 Simply Generated Families

A wide class of planar trees falls under the notion of *simply generated families* (S.G.F.) introduced by *Meir* and *Moon* [41 ]: The trees in a S.G.F. are planted plane trees with associated weights defined by the formal equation

$$S = \{o\} \times (\{\varepsilon\} \mathbin{\dot\cup} c_1 \cdot S \mathbin{\dot\cup} c_2 \cdot S^2 \mathbin{\dot\cup} c_3 \cdot S^3 \mathbin{\dot\cup} \ldots), \tag{2.35}$$

where the numbers $c_i \geq 0$ indicate weight factors corresponding to the outdegrees of the nodes of the given tree. In other words, the o.g.f. will fulfill the functional equation

$$S(z) = z \cdot \varphi(S(z)), \tag{2.36}$$

where $\varphi(t) = 1+c_1t+c_2t^2+ \ldots$ .

Example:

1) $c_1 = c_2 = \ldots = 1 \rightarrow P$
2) $c_1 = 0,\ c_2 = 1,\ c_3 = \ldots = 0 \rightarrow B_0$
3) $c_1 = c_2 = 1,\ c_3 = \ldots = 0 \rightarrow M$
4) $c_1 = 2,\ c_2 = 1,\ c_3 = \ldots = 0 \rightarrow \bar{B}$,

   where $\bar{B}$ consists of the trees in $B$ where all leaves are neglected.

From (2.35) we get

$$s_n = |S_n| \neq 0 \Rightarrow n \equiv 1 (\mathrm{mod}\ d), \text{ where } d = \gcd\{i : c_i > 0\}. \tag{2.37}$$

By the Lagrange Inversion Formula we have

$$s_n = \frac{1}{n} \cdot [z^{n-1}]\ \varphi(z)^n. \tag{2.38}$$

We will see later on, that a number of results on special families of planted plane trees may be generalized to trees falling under this concept.

## 3. SOME REMARKS ON ASYMPTOTIC ENUMERATION

As we have indicated in the Introduction, asymptotic estimates are of great importance in the applications. From the methodological point of view there are two principles to be mentioned (The reader should compare the excellent article [19] by *Flajolet* and *Odlyzko* for more detailed information):

*1) Direct Asymptotics*

Using tools like Stirling's approximation formula or the Euler McLaurin summation formula explicit enumeration formulae are evaluated asymptotically.

Example:

We had $b_n = \frac{1}{n+1}\binom{2n}{n}$. Using Stirling's formula we find directly

$$b_n \sim \frac{1}{\sqrt{\pi}} \cdot 4^n \cdot n^{-3/2} . \tag{3.1}$$

The applicability of direct methods is limited by the facts that explicit formulae must be available and must not be too complicated.

*2) Indirect Asymptotics*

This is the more important principle for practical purposes. The basic idea is that the asymptotic behaviour of sequences is largely determined by the analytic behaviour of (well suited) generating functions, namely location and nature of its singularities, and that the latter information may be gained without having explicit knowledge of the function or its Taylor coefficients.

Let us assume that the sequence $(f_n)$ has the o.g.f. $F(z)$ with radius of convergence $R=1$ and $z_0=1$ is the unique singularity on the circle $|z|=1$. Let us further assume that

$$F(z) \sim c_1(1-z)^{\alpha_1}+c_2(1-z)^{\alpha_2}+\ldots,\ \alpha_1<\alpha_2<\ldots,\ \text{for } z\to 1. \tag{3.2}$$

Then we would like to conclude that

$$f_n = [z^n]\,F(z) \sim c_1 \cdot (-1)^n \binom{\alpha_1}{n} + c_2(-1)^n \binom{\alpha_2}{n} + \ldots,\ n\to\infty . \tag{3.3}$$

More generally speaking an asymptotic expansion of the type

$$F(z)=h_0(z)+h_1(z)+\ldots+h_k(z)+O(g(z)),\ z\to 1 \tag{3.4}$$

$$\text{with } h_0(z) >> h_1(z) >> \ldots >> h_k(z) >> g(z),$$

should translate into

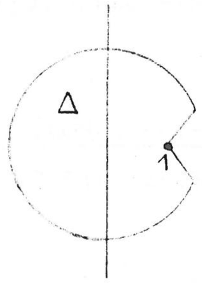
$$f_n = [z^n]\,F(z) = [z^n]\,(h_0(z)+\ldots+h_k(z)) + O([z^n]\,g(z)),\quad n \to \infty . \tag{3.5}$$

It turns out that a transfer result of the desired form is valid under two assumptions:

i) The local expansion (3.4) must hold in a certain region about $z_0=1$ in the complex plane.

ii) The functions $h_i(z)$ and $g(z)$ should belong to a certain asymptotic scale.

Let e.g. $\Delta$ denote a circle with radius $1+\varepsilon$ indented at $z=1$:




Let furthermore $g(z)$ be of the form

$$g(z) = (1-z)^{\alpha} \cdot L\left(\frac{1}{1-z}\right),\quad \alpha \notin \{0,1,2,\ldots\} \tag{3.6}$$

where $L(z)$ is of slow variation towards infinity (compare [19]; $L(z)=\log(z)$ is a characteristic example).

*Then the asymptotic relations*

$$f(z)=O(g(z)),\ f(z)=o(g(z)) \text{ resp. } f(z) \sim g(z) \tag{3.7}$$
$$\text{for } z \to 1,\ z \in \Delta \setminus \{1\}$$

*translate into*

$$f_n = O\left(\frac{n^{-\alpha-1}}{\Gamma(-\alpha)} \cdot L(n)\right),\ f_n = o\left(\frac{n^{-\alpha-1}}{\Gamma(-\alpha)} \cdot L(n)\right) \text{ resp. } f_n \sim \frac{n^{-\alpha-1}}{\Gamma(-\alpha)} \cdot L(n) \tag{3.8}$$
$$\text{for } n \to \infty .$$

The term $\frac{n^{-\alpha-1}}{\Gamma(-\alpha)}$ is the first term in the asymptotic expansion

$$[z^n](1-z)^{\alpha} = (-1)^n \binom{\alpha}{n} \sim \frac{n^{-\alpha-1}}{\Gamma(-\alpha)} \left(1 + \sum_{k\geq 1} \frac{e_k^{(\alpha)}}{n^k}\right),\quad \alpha \notin \{0,1,2,\ldots.\} \tag{3.9}$$

$$\text{where } e_k^{(\alpha)} = \sum_{j=k}^{2k} (-1)^j \lambda_{k,j} (\alpha+1)(\alpha+2)\ldots(\alpha+j),$$

$$\text{with } \sum_{k,j} \lambda_{n,j} u^k t^j = e^t (1+ut)^{-1-1/u} .$$

<u>Example:</u> Let $f(z) \sim (1-z)^{1/2} \cdot \log\left(\frac{1}{1-z}\right)$ for $z \to 1$ in $\Delta\setminus\{1\}$.

Then $f_n \sim -\frac{1}{2\sqrt{\pi}}\, n^{-3/2} \cdot \log n.$

The above transfer result is also applicable to functions having more than one, but finitely many singularities of the mentioned scale on the circle of convergence: in this instance the contributions of all poles must be added up. If the radius of convergence R differs from 1 the simple substitution $z \to \frac{z}{R}$ establishes the situation from above.

Examples 1) Let $f(z)$ have the unique singularity $z=q$ nearest to the origin and

$f(z)=f(q)-a\cdot(q-z)^{1/2}+b(q-z)+O((q-z)^{3/2})$ for $z \to q$, a,b some constants. (3.10)

Then

$$f_n = \frac{a}{2\sqrt{\pi}} \cdot q^{-n+1/2} \cdot n^{-3/2}\left(1+O\left(\frac{1}{n}\right)\right) \text{ for } n \to \infty. \qquad (3.11)$$

If we consider e.g. the o.g.f. of Motzkin trees, we have (compare (2.32))

$$M(z) = \frac{1-z-\sqrt{1-2z-3z^2}}{2z} = \frac{1-z-\sqrt{3(1+z)(\frac{1}{3}-z)}}{2z}$$

so that $q=\frac{1}{3}$ is the dominating singularity.

Since

$$M(z) = 1 - 3\sqrt{\frac{1}{3}-z} + \ldots$$

we have

$$m_n \sim \frac{1}{2} \cdot \sqrt{\frac{3}{\pi}} \cdot 3^n \cdot n^{-3/2} \qquad (3.12)$$

for the Motzkin numbers.

2) Let $\mathcal{S}$ be a simply generated family (compare Section 2.8) with o.g.f. $S(z)$

$$S(z) = z\cdot\varphi(S(z)),$$
$$\varphi(t) = 1+c_1t+c_2t^2+\ldots, \quad c_i \geqq 0.$$

If (i) $\varphi(t)$ has radius of convergence $R>0$.

(ii) There exists $\tau$ with $0<\tau<R$, such that $\tau\cdot\varphi'(\tau)=\varphi(\tau)$ (3.13)

and

(iii) $d := \gcd\{i : c_i > 0\}$:

Then $S(z)$ has radius of convergence $q = \frac{1}{\varphi'(\tau)}$ and d singularities

$$q_k = q\cdot e^{2k\pi i/d}, \quad k=0,1,\ldots,d-1$$

on the circle $|z|=q$. Moreover the local behaviour of $S(z)$ near the singularities is of type (3.10) and

$$s_n = \begin{cases} \frac{d}{2} \cdot \sqrt{\frac{2\tau}{\pi\cdot\varphi''(\tau)}} \cdot q^{-n-1/2}\, n^{-3/2}\left(1+O\left(\frac{1}{n}\right)\right) & \text{for } n\to\infty, n\equiv 1 \pmod d \\ 0 & \text{otherwise} \end{cases} \qquad (3.14)$$

(Compare *Meir/Moon* [41].)

We finally mention that the above transfer method is similar but not identical with *Darboux's* classical *method* (compare [9]):

Instead of the knowledge of the local behaviour of a function $f(z)$ in a sufficiently large region about the singularity, Darboux's method uses smoothness conditions on $f(z)$ like the following:

If $f(z)$ is analytic in $|z|<1$ and k-times continuously differentiable on $|z|=1$ then

$$f_n = o(n^{-k}).$$

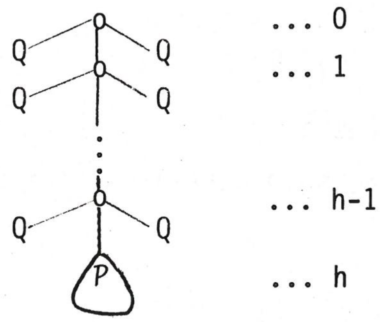
## 4. THE AVERAGE CONTOUR OF PLANE TREES

The average case analysis of several important algorithms is in close connection with questions concerning the *average shape* of certain families of planar trees. In the following sections we give a sketch of some important results in this area.

### 4.1 The Average Level of Nodes

We ask for the average level of a node of a tree in $P_n$, where all trees in $P_n$ are assumed to be equally likely (compare [22]).

Let $Q = \bigcup_{k\geq 0} P^k$ and $C_h$ denote a chain of h elements. Then the family $C_h \times Q^{2h} \times P$ will contain as many copies of each tree in $P$ as there are nodes at level h:




Thus the total number $c(n,h)$ of nodes at level h of all trees in $P_n$ is given by

$$c(n,h) = [z^n]\, z^h \cdot \frac{P(z)}{(1-P(z))^{2h}} \quad . \tag{4.1}$$

With (2.7) we find

$$c(n,h) = \frac{2h+1}{n+h} \binom{2n-2}{n-1-h}, \tag{4.2}$$

and, furthermore, by the Lagrange Inversion formula

$$\sum_h h \cdot c(n,h) = [z^{n-1}] \frac{P}{(1-2P)^2} = 2^{2n-3} - \frac{1}{2}\binom{2n-2}{n-1} .$$

The last expression yields for *the average level of a node in* $P_n$

$$\bar{l}_n = \frac{2^{2n-3}}{\binom{2n-2}{n-1}} - \frac{1}{2} \sim \frac{\sqrt{\pi n}}{2} - \frac{1}{2} + O(n^{-1/2}),\ n\to\infty \tag{4.3}$$

## 4.2 The Average Height of Trees

We start with the problem of *the average height of a tree in* $P_n$ and give a brief sketch of the o.g.f. approach to this problem by *DeBruijn, Knuth* and *Rice* [5]:

Let $A_{n,h}$ be the number of trees in $P_n$ with height less than h and $A_h(z) = \sum_n A_{n,h} z^n$. Then, from (2.6), we have

$$A_{h+1}(z) = \frac{z}{1-A_h(z)}\ ,\ h \geqq 0,\ A_0(z)=0 \tag{4.4}$$

Thus we get from the theory of continued fractions

$$A_h(z) = z \cdot d_h(z)/d_{h+1}(z)$$

with

$$d_h(z) := \varepsilon^{-1}\left(\left(\frac{1+\varepsilon}{2}\right)^h - \left(\frac{1-\varepsilon}{2}\right)^h\right),\ \varepsilon = \sqrt{1-4z}\ . \tag{4.5}$$

Let $B_{n,h} = |P_n| - A_{n,h}$ be the number of trees in $P_n$ with height $\geqq h$. Then

$$B_{n,h} = [z^n]\,(P(z)-A_h(z)),$$

where we remark that $P(z) = \frac{1-\varepsilon}{2}$ .

With the substitution $z = \frac{u}{(1+u)^2}$ , resp. $u = \frac{1-\varepsilon}{1+\varepsilon}$ ,

$$B_{n,h} = [u^n]\,(1-u)^2(1+u)^{2n-2} \cdot \frac{u^{h+1}}{1-u^{h+1}}\ ,$$

so that the average height is

$$\overline{h_n} = \frac{1}{|P_n|} \sum_h h(A_{n,h+1}-A_{n,h}) = \frac{1}{|P_n|} \cdot \sum_{h\geqq 1} B_{nh}$$

$$= -1 + \frac{1}{|P_n|}\,[u^n]\,(1-u)^2(1+u)^{2n-2} \cdot \sum_{h\geqq 1} \frac{u^h}{1-u^h}\ . \tag{4.6}$$

Now

$$\sum_{h\geqq 1} \frac{u^h}{1-u^h} = \sum_{k\geqq 1} d(k)u^k\ , \tag{4.7}$$

where d(k) is the *divisor function*, and we get the *explicit formula*

$$\overline{h_n} = -1 + \frac{1}{|P_n|} \cdot \sum_{k\geqq 1} d(k)\left(\binom{2n}{n+1-k} - 2\binom{2n}{n-k} + \binom{2n}{n-1-k}\right). \tag{4.8}$$

An *asymptotic expansion* of (4.8) may be given using the approximation (compare [28])

$$\binom{2n}{n+a-k}/\binom{2n}{n} = e^{-k^2/n} \cdot f_a(n,k)$$

where e.g. (4.9)

$$f_a(n,k) = 1 - \frac{a^2}{n} + \left(\frac{2a}{n} - \frac{2a^3+a}{n^2}\right)k + \frac{4a^2+1}{2n^2}k^2 + \frac{4a^3+5a}{3n^3}k^3$$

$$- \frac{1}{6n^3}k^4 - \frac{a}{3n^4}k^5 + O(n^{-2+\delta}),\ \delta > 0.$$

Furthermore we need an approximation of

$$g_m(n) = \sum_{k \geqq 1} d(k) \cdot k^m \cdot e^{-k^2/n} :$$

Using the *Mellin Transform*

$$f^*(z) = \int_0^\infty f(t)t^{z-1}\,dt \leftrightarrow f(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} f^*(z)t^{-z}dz$$

we have especially

$$\Gamma(z) = \int_0^\infty e^{-t}\,t^{z-1}\,dt \leftrightarrow e^{-t} = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \Gamma(z)t^{-z}dz \quad (c>0,\ \mathrm{Re}\ t>0)$$

so that

$$g_m(n) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} n^z \cdot \Gamma(z)\left(\sum_{k \geqq 1}\frac{d(k)}{k^{2z-m}}\right)dz. \tag{4.10}$$

In other words: the inverse Mellin transform translates our problem into the study of a Dirichlet series! Now

$$\sum_{k \geqq 1} d(k) \cdot k^{-s} = \zeta^2(s),\ \mathrm{Re}\ s > 1.$$

Shifting the contour in (4.10) to the left of $c = \frac{m+1}{2}$ we get

$$g_m(n) \sim \sum \mathrm{Res}(n^z \cdot \Gamma(z)\zeta^2(2z-m); z_0) \tag{4.11}$$

where the sum is to be taken over all poles $z_0$ with Re $z_0 \leqq \frac{m+1}{2}$ .

With this technique we finally get

$$\overline{h_n} = \sqrt{\pi n} - \frac{3}{2} + O(n^{-1/2+\delta}),\ \delta > 0. \tag{4.12}$$

Comparing (4.12) with (4.3) we have

$$\overline{h_n} = 2\,\overline{l_n} - \frac{1}{2} + O(n^{-1/2+\delta}) \tag{4.13}$$

and the question arises whether there is a direct combinatorial estimate of $\overline{h_n}$ via $2\,\overline{l_n}$.

In [12] *Dershowitz* and *Zaks* give an estimate of this type using the Cycle

Lemma (compare Section 2.5):

As in the proof of (2.28) we associate with each tree in $P_{n+1}$ a 1-dominating sequence of n+1 boxes and n balls, the latter corresponding to a positive random walk from (0,0) to (2n+1,1).

From the Cycle Lemma we know that all cyclic shifts of the above walks produce *all* walks from (0,0) to (2n+1,1) exactly once.

The level l of a node may be measured at the bottom of the corresponding upward step in the original path. After the cyclic shift the same step will have (signed bottom) level

$$d = \begin{cases} l-s & \text{if the step was right of the shift cut} \\ l-s+1 & \text{if it was left,} \end{cases} \tag{4.14}$$

where s is the level of the cut in the original path.

Let a denote the minimal (signed bottom) level in the shifted path, then

$$a = \begin{cases} -s = 0 & \text{for the identity shift} \\ 1-s & \text{otherwise .} \end{cases} \tag{4.15}$$

Therefore

$$d = \begin{cases} l+a-1 \\ l+a \end{cases}$$

in (4.14) for nonidentical shifts and d=l for the identity. Thus we get for the means

$$\bar{d} - \bar{a} \leq \bar{l} \leq \bar{d} - \bar{a} + 1 . \tag{4.16}$$

Let h denote the height of the tree, i.e. the maximal (bottom) level of an upward step in the original positive random walk. Let furthermore z denote the maximal (signed bottom) level of an upward step in the shifted walk. In the same way as above we find

$$\bar{z} - \bar{a} \leq \bar{h} \leq \bar{z} - \bar{a} + 1 . \tag{4.17}$$

Consider now together with each positive walk w its reverse $\tilde{w}$. Then it follows immediately that

$$\bar{d} = 0 \quad \text{and} \quad \bar{z} = -\bar{a} . \tag{4.18}$$

Combining (4.16), (4.17) and (4.18) we have

$$2\bar{l} - 2 \leq \bar{h} \leq 2\bar{l} +1,$$

so that

$$\bar{h} \sim \sqrt{\pi n} + O(1) \tag{4.20}$$

is established.

*Some further important results* concerning the height of planted plane trees are:

1)(*Kemp*, [26]) The average height of a tree in $P_n$ with root degree r is asymptotic to

$$\sqrt{\pi n} - \frac{r}{2} + O(n^{-1/2+\delta}), \quad \delta > 0 . \tag{4.21}$$

2)(*Kemp*, [27]) The average height of a tree in $P_{n,m}$ equals

$$\bar{h}_{n,m} = 1 + \frac{1}{|P_{n,m}|} [f_1(m,n) - 2f_0(m,n) + f_{-1}(m,n)]$$

where

$$f_a(m,n) = \sum_{\lambda \geq 1} \sum_{d|\lambda} \binom{n-2-\lambda}{m+d+a-1}\binom{n-2+\lambda}{m-d-a-1}. \tag{4.22}$$

For $m = \rho n$, $0 < \rho < 1$, and $n \to \infty$

$$\bar{h}_{n,m} = \sqrt{\pi \cdot \frac{1-\rho}{\rho}} \cdot \sqrt{n} + \frac{1}{2} - \frac{1}{\rho} + O(n^{-1/2+\delta}). \tag{4.23}$$

3)(*Prodinger*, [49]) The average height of the d-th highest leaf of a tree in $P_n$ fullfills

$$\bar{h}_n^{(d)} \sim \sqrt{\pi n} - \frac{3}{2} - \sum_{s=0}^{d-2} \frac{2}{(s+1)3^{s+1}} [x^s] \frac{1}{(1-x)^2} \left(\frac{1+x}{1-x/3}\right)^{s+1} + O(n^{-1/2+\delta}) \tag{4.24}$$

so that e.g.

$$\bar{h}_n - \bar{h}_n^{(2)} \to 2/3, \quad h_n - h_n^{(3)} \to 32/27. \tag{4.25}$$

4)(*Kemp*, [29]) The average number of nodes at the maximum level for trees in $P_n$ tends to 2 for $n \to \infty$. If all trees in $P_n$ with height k are equally likely the average number of nodes at level k is

$$\sim 4 \cdot \frac{n+1}{k+1} \cdot \sin^2\left(\frac{\pi}{k+1}\right) - \frac{6}{k+1} + O\left(\frac{1}{n}\right), \quad n \to \infty, \text{ fixed } k. \tag{4.26}$$

3) and 4) make use of the following combinatorial observation:
Let $T_{n,k,r}$ be the number of trees in $P_n$, with height $\leq k$ and root degree r, and let $Q_{n,k,r}$ be the number of trees in $P_n$, with height $=k$ and exactly r nodes at level k. Then

$$Q_{n,k,r} = T_{n+1,k,r+1} - T_{n+1,k,r} + T_{n,k,r-1}; \quad n,k,r > 0. \tag{4.27}$$

(*Kemp*, [29]). In [53] *Strehl* gives two short proofs for (4.27). One of them uses the fact that from the construction of $P$ we have the following o.g.f.'s of continued fraction type:

$$T_k(z,u) = \sum_{n \geq 1, r \geq 0} T_{n,k,r} z^n u^r = \left.\frac{z}{1 - \cfrac{zu}{1 - \cfrac{z}{\ddots\, \cfrac{z}{1-z}}}}\right\} k \tag{4.28}$$

resp.

$$Q_k(z,u)=\sum_{n\geq 1,r\geq 0} Q_{n,k,r}\, z^n u^r = \left.\cfrac{z}{1-\cfrac{z}{\ddots\, \cfrac{z}{1-zu}}}\right\} k \tag{4.29}$$

so that

$$T_k(z,u)=z\cdot\frac{a_{k-1}(z,1)}{a_k(z,u)}\ ,\quad Q_k(z,u)=z\cdot\frac{a_{k-1}(z,u)}{a_k(z,u)}\ ,$$

with

$$\begin{aligned} a_0(z,u) &= 1\\ a_1(z,u) &= 1-zu\\ a_k(z,u)&=a_{k-1}(z,u)-z\cdot a_{k-2}(z,u),\ k\geq 2\ . \end{aligned} \tag{4.30}$$

Identity (4.27) follows now from

$$(1-u)a_k(z,u)+zu\ a_{k-1}(z,u) = (1-u+zu^2)a_{k-1}(z,1). \tag{4.31}$$

5)(*Kirschenhofer, Prodinger* [34]) Let $h_k(t)$ be the maximal number of nodes of degree k in a chain connecting the root with a leaf. The average of $h_k$ in $\mathcal{P}_n$ fulfills

$$\overline{h_{k,n}} \sim \frac{k}{2^{k+1}}\sqrt{\pi n}\ ,\quad n\to\infty\ . \tag{4.32}$$

In the second part of this section we present an outline of the analysis of *the average height of binary trees* and other simple families, following the pioneering paper by *Flajolet* and *Odlyzko* [19].

Remembering that the average "left-sided" height $\bar{h}_L$ of a tree in $\mathcal{B}_n$ equals via the Rotation Correspondence (Section 2.4) the average height of a tree in $\mathcal{P}_{n+1}$, we might guess

$$\bar{h}(\mathcal{B}_n)\sim 2\bar{h}_L(\mathcal{B}_n)\sim 2\cdot\sqrt{\pi n}\ . \tag{4.33}$$

Nevertheless a proof of (4.33) is by no means trivial. The ogf-approach starts with $A_h(z)$, where $[z^n]A_h(z)$ is the number of trees in $\mathcal{B}_n$ with height $<h$. From the construction of $\mathcal{B}$ (2.10) we get

$$\begin{aligned} A_{h+1}(z) &= 1+z\cdot(A_h(z))^2,\ h\geq 0\\ A_0(z) &= 0. \end{aligned} \tag{4.34}$$

Now $H(z)$ with $[z^n]H(z)=\sum_{t\in\mathcal{B}_n} h(t)$ is given by

$$H(z)=\sum_{h\geq 1}(B(z)-A_h(z)).$$

If we set

$$f_h(z) := \frac{B(z)-A_h(z)}{2B(z)} , \tag{4.35}$$

so that

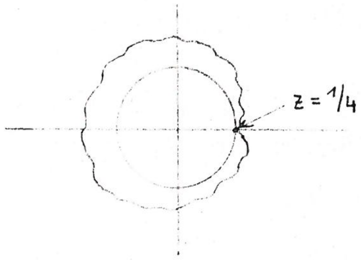
$$H(z) = 2B(z) \cdot \sum_{h \geqq 1} f_h(z), \tag{4.36}$$

we have the recurrence

$$\begin{aligned} &f_0(z) = \frac{1}{2} \\ &f_{h+1}(z) = (1-\varepsilon) f_h(z)(1-f_h(z)), \; \varepsilon = \sqrt{1-4z}, \; h \geqq 0. \end{aligned} \tag{4.37}$$

The main part of the analysis is devoted to the study of the convergence of $f_h(z)$ and split up into several parts (we have to omit the proofs here):

i) $f_h(z)$ converges geometrically and uniformly in a region of the type




For all $\eta > 0$ there exists $\lambda > 1/4$ such that $\sum_h f_h$ is analytic for $|z| < \lambda$, $|\operatorname{Arg} z| > \eta$.

ii) In order to study the local behaviour of $\sum f_h$ about $z=1/4$ recurrence (4.37) is transformed according to an idea of *De Bruijn* [4] by taking inverses and multiplying by $(1-\varepsilon)^{h+1}$:

$$\frac{(1-\varepsilon)^{h+1}}{f_{h+1}} = \frac{(1-\varepsilon)^h}{f_h} + (1-\varepsilon)^h + \frac{f_h}{1-f_h} (1-\varepsilon)^h ,$$

so that, after summing up, we get the *alternative recurrence*

$$\frac{(1-\varepsilon)^h}{f_h} = \frac{1-(1-\varepsilon)^h}{\varepsilon} + 2 + \sum_{j<h} \frac{f_j}{1-f_j} (1-\varepsilon)^j . \tag{4.38}$$

Using (4.38) it can be shown (by a number of delicate estimates) that $\sum_h f_h(z)$ is analytic inside of a circle $\Delta$ of radius $\frac{1}{4} + \delta_1$ indented at $z=1/4$ (compare Section 3) and is approximated by $\sum_h \frac{\varepsilon(1-\varepsilon)^h}{1-(1-\varepsilon)^h}$. With $1-\varepsilon(z)=e^{-u}$

$$\sum_h f_h(z) \approx \frac{1-e^{-u}}{u} \cdot \sum_j u \cdot \frac{e^{-ju}}{1-e^{-ju}} . \tag{4.39}$$

The right-hand side may be compared with the integral

$$\int_u^\infty \frac{e^{-x}}{1-e^{-x}} \, dx$$

yielding finally

$$H(z) = -2\log(1-4z)+K+O(|1-4z|)^{1/4-\delta} \tag{4.40}$$

for z in $\Delta$ from above. This local expansion transfers into (compare Section 3):

$$\sum_{t\in\mathcal{B}_n} h(t) = 2\cdot 4^n \cdot \frac{1}{n}\,(1+O(n^{-1/4+\delta})) \tag{4.41}$$

so that we get the desired result

$$\bar{h}_{\mathcal{B}_n} = 2\cdot\sqrt{\pi n} + O(n^{1/4+\delta}). \tag{4.42}$$

The same technique allows to establish the average height of the n-node trees in a simply generated family $\mathcal{S}$ : With the notions of Section 2.8 one gets

$$\bar{h}_{\mathcal{S}_n} \sim \sqrt{\frac{2\pi}{\varphi(\tau)\varphi''(\tau)}}\cdot\varphi'(\tau)\cdot\sqrt{n}\,,\quad n\to\infty\,. \tag{4.43}$$

This result covers the families $\mathcal{P}$, $\mathcal{B}$, $\mathcal{M}$, $\mathcal{T}$ and even the family $\mathcal{L}$ of labelled non-planar trees (Section 7), where the result was proved earlier by *Renyi* and *Szekeres* [51] using probabilistic arguments.

## 4.3 The Average Height of Specified Endnodes

In order to analyze the average contour of planted plane trees more accurately it seems convenient to study the average height of specified endnodes. Let us assume that the n+1 endnodes of a tree in $\mathcal{B}_n$ are enumerated by 0,1,...,n from the left to the right. Then we denote by $\alpha_{\mathcal{B}}(n,j)$ the *average height of leaf number "j"* in $\mathcal{B}_n$. In [44] *Moon* has proved that

$$\alpha_{\mathcal{B}}(n,\tfrac{n}{2}) \sim \frac{4}{\sqrt{\pi}}\cdot\sqrt{n},\quad n\to\infty \tag{4.44}$$

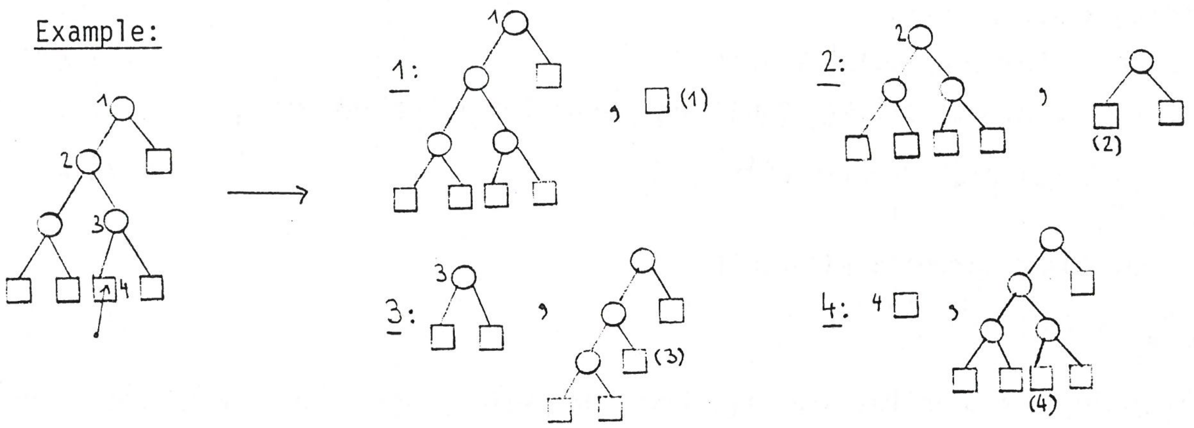
In fact the following explicit formula holds (compare the author's papers [30], [31]):

$$\alpha_{\mathcal{B}}(n,j) = \frac{4(n+1)(2n+1)}{n+2}\cdot\frac{\binom{n}{j}^2}{\binom{2n+2}{2j+1}} - 1\,. \tag{4.45}$$

The proof may be given by ogf-techniques or by a more direct combinatorial reasoning, which we present here in short:

Consider for each tree t in $\mathcal{B}_n$ the (unique) path connecting the root with leaf number "j". For each vertex in this path we form a pair of trees $(t_1,t_2)$ in that way, that $t_1$ is the subtree of t whose root is the vertex in consideration, and $t_2$ is the remaining tree where the vertex in consideration is substituted by a leaf. Altogether we will get $h_j(t)+1$ pairs of binary trees, where $h_j(t)$ is the height of leaf number "j" in t.

Example:



Thus $\sum_{t\in \mathcal{B}_n} (h_j(t)+1)$ may be computed by enumerating the pairs of trees created above:

Each pair is of the form $(t_1,t_2)$, $t_1 \in \mathcal{B}_\mu$, $t_2 \in \mathcal{B}_{n-\mu}$. Let us assume w.l.o.g that $j \le \frac{n}{2}$ :

Case 1) $0 \le \mu \le j-1$: Then the given pair will occur $\mu+1$-times, since there are $\mu+1$ possibilities to adjoin $t_1$ to $t_2$, such that endnode number "j" of this larger tree is one of the endnodes of $t_1$.

Case 2) $j \le \mu \le n-j$: There are $j+1$ possibilities.

Case 3) $n-j+1 \le \mu$: There are $n-\mu+1$ possibilities.

Altogether we have

$$\sum_{t\in \mathcal{B}_n} (h_j(t)+1) = \sum_{\mu=0}^{n} (1+\min\{\mu,j,n-\mu\} b_\mu \cdot b_{n-\mu}) =: S. \tag{4.46}$$

A short calculation shows

$$S = S_1 - S_2 + (j+1)b_{n+1},$$

where

$$\begin{aligned} S_1 &= (-1)^{n+1}\, 4^{n+1} \sum_{\mu=0}^{n+1-j} \binom{1/2}{\mu}\binom{-1/2}{n+1-\mu}, \\ S_2 &= 2(j+1)4^{n+1}(-1)^n \sum_{\mu=1}^{j} \binom{1/2}{\mu}\binom{1/2}{n+2-\mu}. \end{aligned} \tag{4.47}$$

The evaluation of partial Vandermonde convolutions as in $S_1$ and $S_2$ can be performed via the following pair of identities by

*E. Sparre Andersen:*

$$\begin{aligned} \sum_{i=0}^{k} \binom{a}{i}\binom{-a}{n-i} &= \frac{n-k}{n} \binom{a-1}{k}\binom{-a}{n-k}, \\ \sum_{i=0}^{k} \binom{a}{i}\binom{1-a}{n-i} &= \frac{(n-1)(1-a)-k}{n(n-1)} \binom{a-1}{k}\binom{-a}{n-k-1}, \quad 0\le k\le n,\ n\le 1. \end{aligned} \tag{4.48}$$

Andersen's proof is by induction. A more direct proof is by expressing the sums in terms of *hypergeometric series* (*D. Foata*, private communication):

$$\sum_{i=0}^{k}\binom{a}{i}\binom{-a}{n-i} = -\sum_{i=k+1}^{n}\binom{a}{i}\binom{-a}{n-i} \quad \text{by Vandermonde}$$

$$= -\sum_{j=0}^{n-k-1}\binom{a}{j+k+1}\binom{-a}{n-j-k-1}$$

$$= (-1)^{n+1}\frac{(a)_n}{n!}\cdot\frac{(-a)_{k+1}}{(k+1)!}\cdot\frac{(-n)_{k+1}}{(-a-n+1)_{k+1}}\cdot\sum_{j=0}^{n-k-1}\frac{(-n+k+1)_j\cdot(-a+k+1)_j}{(k+2)_j\cdot(2-a-n+k)_j},$$

where $(a)_0 = 1$, $(a)_n = a(a+1)\ldots(a+n-1)$.

The sum equals the hypergeometric series

$$ {}_3F_2\left[\begin{matrix}-n+k+1,\ -a+k+1,\ 1\\ k+2,\ 2-a-n+k\end{matrix};1\right], \tag{4.49}$$

where we adopt the standard notation

$$ {}_pF_q\left[\begin{matrix}a_1,\ldots,a_p\\ b_1,\ldots,b_q\end{matrix};x\right] = \sum_{n\geq 0}\frac{(a_1)_n\cdots(a_p)_n}{(b_1)_n\cdots(b_q)_n}\cdot\frac{x^n}{n!}.$$

In order to simplify (4.49) we use the *Pfaff-Saalschütz formula* (compare [24])

$$ {}_3F_2\left[\begin{matrix}-n,a,b\\ c,-n+a+b+1-c\end{matrix};1\right] = \frac{(c-a)_n(c-b)_n}{(c)_n(c-a-b)_n} \tag{4.50}$$

and get the right hand side of (4.48) immediately.

Formula (4.45) allows direct asymptotic expansions. So we get e.g. for the central region of the tree:

$$\alpha_{\mathcal{B}}(n,\rho_n) = \sqrt{\rho(1-\rho)}\cdot\frac{8}{\sqrt{\pi}}\cdot\sqrt{n} - 1 + O(n^{-1/2}),\quad 0<\rho<1,\quad n\to\infty. \tag{4.51}$$

To derive *similar results for trees in* more complicated *simply generated families* is a difficult task, since no simple explicit formulas like (4.45) are available. The same enumerative technique as with family $\mathcal{B}$ shows for a general simple family $\mathcal{S}$ that the ogf $H(z,u)=\sum_{n,j} z^n u^j \sum_{t\in \mathcal{S}_n} h_j(t)$ is given by

$$H(z,u) = \frac{u}{z}\cdot\left(\frac{S(z)-S(z,u)}{1-u}\right)^2 \tag{4.52}$$

where $[z^n u^j]S(z,u)$ is the number of trees in $\mathcal{S}_n$ with exactly j leaves, and $S(z)=S(z,1)$. The problems is now, to find the asymptotic behaviour of the coefficients of $H(z,u)$, where $S(z)$ resp. $S(z,u)$ are only given implicitly (compare equation 2.36).

For fixed j and $n\to\infty$ it is easy to show that

$$\alpha_{\mathcal{S}}(j) = \lim_{n\to\infty}\alpha_{\mathcal{S}}(n,j)$$

exists, and to analyze its behaviour for j getting large by a transfer technique (Section 3).

The more interesting instance (as in 4.51 above) is $j=\rho n$, $0<\rho<1/\varphi(\tau)$, $n\to\infty$. ($\varphi(\tau)^{-1}\cdot n$ is the expected value of the number of leaves of a tree in $S_n$.)
Let $\rho = \frac{p}{r} \in \mathbb{Q}$, $\gcd(p,r)=1$, $0<p<r$. We need the ogf

$$H_{p,r}(x) = \sum_m ([z^{rm}u^{pm}]H(z,u))x^{rm} \tag{4.53}$$

of a *"diagonal"* of $H(z,u)$. The main idea is to use the residue calculus and express

$$H_{p,r}(x) = \frac{1}{2\pi i} \int_{C(x)} H(\frac{x}{s^p}, s^r) \frac{ds}{s}, \tag{4.54}$$

where $C(x)$ is an appropriate contour separating those singularities in s of $H(\frac{x}{s^p}, s^r)$ (x fixed) that tend to 0 for $x\to 0$ from the other ones.

The most difficult part of the analysis is to find a local expansion of the integrand as a function in s that holds uniformly in x in a certain region. As a consequence we find a local expansion of $H_{p,r}(x)$ and, via transfer techniques, an asymptotic expansion of the coefficients.

Following this idea the author could prove [32], [33]:

$$\alpha_S(n,\rho_n) \sim \sqrt{n} \cdot \sqrt{\rho\left(\frac{1}{\varphi(\tau)} - \rho\right)} \cdot 8 \cdot \varphi'(\tau) \cdot \sqrt{\frac{\varphi(\tau)}{2\pi\varphi''(\tau)}}$$
$$\text{for } 0<\rho<\frac{1}{\varphi(\tau)}, \quad n\to\infty, \tag{4.55}$$

where $S(z) = z\cdot\varphi(S(z))$ and $\tau\varphi'(\tau) = \varphi(\tau)$.

Example: For $P$ we have $\varphi(t) = (1-t)^{-1}$, $\tau=1/2$, so that

$$\alpha_P(n,\rho_n) \sim \frac{8}{\sqrt{\pi}} \sqrt{\rho(\tfrac{1}{2}-\rho)} \cdot \sqrt{n}, \quad n\to\infty .$$

## 4.4 Level Number Sequences of Binary Trees

The level number sequence $\mathrm{lns}(t) = (n_0, n_1, n_2, \dots)$ of a tree $t \in B$ is the sequence where $n_i$ counts the number of internal nodes of t at level i. Let $H_n$ denote the set of all different lns of trees in $B_n$ and $H_n = |H_n|$. We are interested in the asymptotics of $H_n$ (*Flajolet, Prodinger* [20]):
The set $H_{n,k}$ of lns of order k has elements of the form $(n_0, n_1, \dots, n_{k-1}, 0, 0, \dots)$ with $n_{k-1} \neq 0$. In other words

(i) $n_0 = 1$

(ii) $1 \le n_j \le 2n_{j-1}$ for all $1 \le j \le k-1$

and (iii) $n_0 + n_1 + \ldots + n_{k-1} = n$

characterizes the elements of $H_{n,k}$, which may, for this reason, be also considered as certain restricted compositions of integers.

A table of the first values of $H_n$ looks like:

| n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $H_n$ | 1 | 1 | 2 | 3 | 5 | 9 | 16 | 28 | 50 | 89 |

It is easily seen that

$$F_n \leqq H_n \leqq 2^{n-1}, \tag{4.56}$$

where $F_n$ is the n-th Fibonacci number. (For the left inequality we count compositions with summands only 1 or 2, for the right one we count all unrestricted compositions of n.)

Let $H_{n,k,j}$ be the number of different lns in $H_{n,k}$ with last non-zero component equal to j (we denote the set by $H_{n,k,j}$) and

$$\begin{aligned} H^{[k]}(q,u) &:= \sum_{n,j\geqq 1} H_{n,k,j}\, q^n u^j, \\ H(q,u) &:= \sum_{k\geqq 1} H^{[k]}(q,u) \end{aligned} \tag{4.57}$$

the corresponding ogf. Then

$$H(q) = \sum_{n\geqq 1} H_n q^n = H(q,1). \tag{4.58}$$

Considering the elements of $H_{n,k,j}$ and adding a new non-zero component $n_{k+1}$ we get lns of order k+1 with last non-zero component $j^* \in \{1,2,\ldots,2j\}$ and total sum $n+j^*$. This means in the ogf $H^{[k]}(q,u)$ the substitution

$$u^j \to uq + (uq)^2 + \ldots + (uq)^{2j} = \frac{uq}{1-uq}\,(1-(uq)^{2j}),$$

so that

$$H^{[k+1]}(q,u) = \frac{uq}{1-uq}\left[H^{[k]}(q,1) - H^{[k]}(q,q^2u^2)\right]$$

$$H^{[0]}(q,u) = qu$$

and

$$H(q,u) = qu + \frac{qu}{1-qu}\,[H(q,1) - H(q,q^2u^2)]. \tag{4.59}$$

This is an equation of the type

$$\Phi(u) = \lambda(u) + \mu(u)\cdot\Phi(\sigma(u)),$$

which has the formal solution

$$\Phi(u) = \sum_{k\geqq 0}\left[\prod_{j=0}^{k-1} \mu(\sigma^{(j)}(u))\right]\lambda(\sigma^{(k)}(u)), \tag{4.60}$$

where $\sigma^{(i)}(u)$ is the i-th iterate of $\sigma$.

Applying (4.60) to (4.59) we get a solution of the form

$$H(q,u) = A(q,u)+B(q,u)\cdot H(q,1),$$

so that

$$H(q)=H(q,1)= \frac{A(q,1)}{1-B(q,1)} = \frac{a(q)}{1-b(q)} . \tag{4.61}$$

$a(q)$ and $b(q)$ are the following beautiful q-series

$$a(q)= \sum_{j\geq 1} (-1)^{j+1}q^{2^{j+1}-2-j}\cdot P_{j-1}(q), b(q)= \sum_{j\geq 1} (-1)^{j+1}q^{2^{j+1}-2-j}\cdot P_j(q) \tag{4.62}$$

where

$$P_i(q) = (1-q)(1-q^3)\ldots(1-q^{2^i-1}).$$

From (4.61) it follows by the transfer method that

$$H_n \sim K\cdot \rho^{-n}, \; n \to \infty , \tag{4.63}$$

where $\rho$ is the smallest positive root of $b(x)=1$,

$$\rho^{-1} = 1{,}794\ 147\ldots$$
$$K = 0{,}254\ 505\ldots$$

The numbers $H_n$ allow also some other interesting interpretations:

1)(*Sloane*, [52] ) $H_n$ is the number of possibilities to write 1 as the sum of n+1 terms of $\{2^{-k}, k\geq 0\}$, where repetitions are allowed and the order is irrelevant.

2)(*Lannes*, see [20]) $H(q)$ is the Poincaré series of the module on Steenrod's algebra.

## 5. DIFFERENT STATISTICS ON TREES: THE DIGITAL SEARCH TREE MODEL

### 5.1 Digital Search Trees, Tries and Patricia Tries

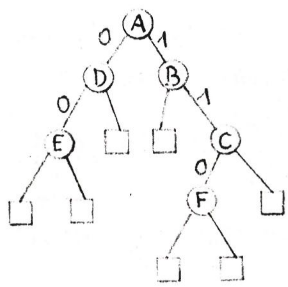
In this section we study data structures which make use of the digital properties of keys. Each record is represented by a key which is assumed to be an (infinitely long) 0,1-sequence, where 0 and 1 may occur with equal probability. In the *digital search tree* (DST, first proposed by *Coffman* and *Eve* [7]) we build up a binary tree which contains the records in its internal nodes. The first record is stored in the root, the following records are stored in the first empty internal node, where the left-right decision is governed by the bits of the keys. For example:

A : 0100...   D : 0011...

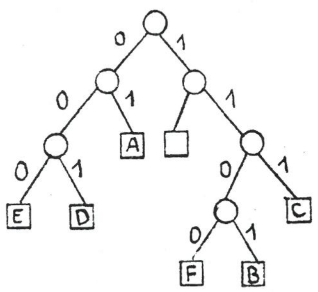
B : 1101...   E : 0000...

C : 1110...   F : 1100...

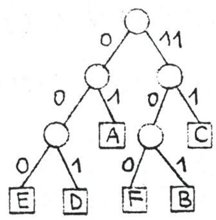
$\longleftrightarrow$



*Digital tries* (from information re*trie*val) follow the same construction principle, but storing the records in the leaves instead of the internal nodes. In other words: the position of a record is determined by the shortest unique prefix of the key. It should be noted that the relative order of the records is irrelevant in this model, but it is relevant, when a digital search tree is constructed. The trie corresponding to the keys from the last example is depicted left below. It is easily seen that this tree may be compressed in order to avoid endnodes with null entry. The corresponding structure is called *Patricia trie* (from "*p*r*a*ctical *a*lgorithm *t*o *r*etrieve *i*nformation *c*oded *i*n *a*lphanumeric") and depicted right below:




Trie




Patricia Trie

## 5.2 The Internal Path Length in Digital Search Trees

In the *average case analysis* of search algorithms for the above mentioned data structures the *path length* is the most important parameter:

The internal (resp. external) path length of a DST (resp. trie or Patricia trie) is the sum of lengths of all paths from the root to an occupied internal (resp. external) node. The average number of nodes examined during a *successful search* in a DST with N records is 1/N times the internal path length incremented by 1, compare *Knuth* [38].

In the following we give a short sketch of the analysis of the expectation and the variance of the internal path length for DST (following *Prodinger*, *Szpankowski* and the author [36]).

We start by setting up a recurrence relation for the probability generating functions $F_N(z)$, where $[z^k]F_N(z)$ is the probability that a DST with N records has path length equal to k:

$$F_{N+1}(z)=z^N\cdot\sum_{k=0}^{N}\binom{N}{k}2^{-N}\cdot F_k(z)F_{N-k}(z);\quad N\geq 0;\quad F_0(z)=1, \qquad (5.1)$$

since $\binom{N}{k}\cdot 2^{-N}$ is the probability that k of N keys start with 0 (and therefore are stored in the left subtree).

Consequently the *expectation* $l_N=F_N'(1)$ of the internal path length fulfills

$$l_{N+1} = N+2^{1-N}\sum_{k=0}^{N}\binom{N}{k}l_k,\ N\geq 0;\ l_0=0. \tag{5.2}$$

In order to solve (5.2) we use the exponential g.f.

$$L(z) = \sum_{N\geq 0} l_N z^N/N!\ .$$

From (5.2) we find the differential functional equation

$$L'(z) = z\cdot e^z+2e^{z/2}\,L(\tfrac{z}{2})\ .$$

After the substitution $\hat{L}(z)=e^zL(-z)$ we find

$$\hat{L}'(z) = z+\hat{L}(z)-2\hat{L}(\tfrac{z}{2}),$$

so that

$$\hat{l}_{N+1}=\hat{l}_N(1-2^{2-N}),\ N\geq 2;\ \hat{l}_2=1,\ \hat{l}_1=\hat{l}_0=0,$$

or

$$\hat{l}_N = Q_{N-2},\ N\geq 2;\ \hat{l}_0=\hat{l}_1=0,$$

$$\text{with } Q_m = \prod_{i=1}^{m}(1-2^{-i})\ . \tag{5.3}$$

Finally we have the explicit solution

$$l_N = \sum_{k=2}^{N}\binom{N}{k}(-1)^kQ_{k-2},\ N\geq 2;\ l_0=l_1=0. \tag{5.4}$$

An asymptotic evaluation of (5.4) is not immediate, since for N getting large we have an alternating sum where the single terms have almost equal size. It is convenient to use the following Lemma from the calculus of finite differences, in order to transform the discrete sum into a complex contour integral:

*Lemma*([46]): Let $C$ be a curve surrounding the points $z=s,s+1,\ldots,N$ in $\mathbb{C}$, $f(z)$ analytic inside $C$ and continuous along $C$. Then

$$\sum_{k\geq s}\binom{N}{k}(-1)^kf(k)= -\frac{1}{2\pi i}\int_C [N;z]f(z)dz, \tag{5.5}$$

with

$$[N;z] = \frac{(-1)^{N-1}\cdot N!}{z(z-1)\ldots(z-N)}\ .$$

So what we need is a complex function $f(z)$ which interpolates the given values $f(s),\ldots,f(N)$. Furthermore $f(z)$ should obey certain growth estimates, that allow to extend the contour of integration $C$ in (5.5) to a large rectangle and to show that the right, upper and left parts of the rectangle tending to $+\infty$ resp. $\pm i\infty$ give negligible small contributions to the integral. Altogether we get an asymptotic expansion

$$\sum_{k\geq s}\binom{N}{k}(-1)^k f(k) = \sum \mathrm{Res}(f(z)\,[N;z]\,;z=z_j)+O(N^c) \tag{5.6}$$

where the sum is over all poles different from $s,s+1,\ldots,N$ with real part $>c$.

In our example we have $f(k)=Q_{k-2}$ and may take

$$f(z)=\frac{Q_\infty}{Q(2^{2-z})}\,,\ \text{with}\ Q(t)=\prod_{i\geq 1}(1-t2^{-i}),\ Q_\infty = Q(1). \tag{5.7}$$

The dominating ($2^{nd}$ order) pole is $z=1$, and there is an infinity of ($1^{st}$ order) poles of same real part at $z = 1+\chi_k=1+\frac{2k\pi i}{\log 2}$, $k\in\mathbb{Z}$, $k\neq 0$. In a similar way we find a ($2^{nd}$ order) pole at $z=0$ and ($1^{st}$ order) poles at $z=\chi_k$, $k\neq 0$. Collecting the residues it turns out that the poles regularly distributed parallel to the imaginary axis give rise to periodic fluctuations of $l_N$:

$$\begin{aligned} l_N &= N\cdot\log_2 N + N\left(\frac{\gamma-1}{\log 2}+\frac{1}{2}-\alpha+\delta_1(\log_2 N)\right)\\ &+ \log_2 N + \frac{2\gamma-1}{2\log 2}+\frac{5}{2}-\alpha+\delta_2(\log_2 N)\\ &+ O(N^{-1+\varepsilon}), \end{aligned} \tag{5.8}$$

where $\gamma$ is Euler's constant, $\alpha = \sum_{n\geq 1}\frac{1}{2^n-1} = 1{,}60669\ldots$, and $\delta_1(x)$ resp. $\delta_2(x)$ are continuous periodic functions of mean zero and amplitude $<10^{-6}$. The Fourier expansion of $\delta_1(x)$ resp. $\delta_2(x)$ follows from the above derivation (residues at $1+\chi_k$, $\chi_k$, $k\neq 0$). E.g.

$$\delta_1(x) = \frac{1}{\log 2}\cdot\sum_{k\neq 0}\Gamma\left(-1-\frac{2k\pi i}{\log 2}\right)e^{2k\pi i x}\,. \tag{5.9}$$

Equation (5.8), although with a less accurate asymptotic expansion, has been established using different methods by *Konheim* and *Newman* [39], *Knuth* [38], and *Flajolet* and *Sedgewick* [21].

Considering the variance

$$V_N = F_N''(1) + F_N'(1) - F_N'(1)^2 \tag{5.10}$$

two main difficulties occur:

We need an accurate asymptotic expansion of the $2^{nd}$ factorial moment $F_N''(1)$, and we need information on the mean of $\delta_1^2(x)$, occurring in $F_N'(1)^2$ in (5.10).

Referringto the first problem we mention that in a similar way as for $l_N$ we find the following recurrence for $s_N=F''_N(1)$:

$$s_{N+1}=N2^{2-N}\sum_{k=0}^{N}\binom{N}{k}l_k+N(N-1)+2^{1-N}\sum_{k=0}^{N}\binom{N}{k}l_k l_{N-k}+2^{1-N}\sum_{k=0}^{N}\binom{N}{k}s_k,\quad N\geq 0. \tag{5.11}$$

Recursion (5.11) may be split up into 3 parts, and solved explicitly. In the solution there occur some very involved alternating sums, like

$$w_N = \sum_{k\geq 5}(-1)^k\binom{N}{k}\hat{w}_k, \tag{5.12}$$

with

$$\hat{w}_k = -Q_{k-2}\sum_{j=1}^{k-1}\frac{1}{2^{j-1}Q_{j-1}}\sum_{i=2}^{j-2}\binom{j}{i}Q_{i-2}Q_{j-i-2}$$

($Q_i$ from (5.3)). In order to find an analytic interpolation of $\hat{w}_k$, one has to find e.g. a function f(z) with

$$f(N+1) = \sum_{k=2}^{N-2}\binom{N}{k}Q_{k-2}Q_{N-k-2}\ . \tag{5.13}$$

Now $Q_k = Q_\infty/Q(2^{-k})$, where $Q(t) = \prod_{n\geq 1}(1-t2^{-n})$.

The main idea is now to use *Euler's product identity* (comp.[1])

$$Q(t)^{-1} = \sum_{n\geq 0} t^n/2^n Q_n, \qquad \text{so that} \tag{5.14}$$

$$f(N+1)=Q_\infty^2\sum_{i,j\geq 0}\left[(2^{-i}+2^{-j})^N-2^{-iN}-2^{-jN}-N\cdot 2^{-i(N-1)-j}-N2^{-i-j(N-1)}\right]2^{i+j}/Q_iQ_j\ .$$

The double sum, which is symmetric in i and j may now be rewritten as $\sum_{i,j\geq 0} = 2\sum_{j\geq i\geq 0}-\sum_{i=j}$.

After the substitution j=i+h we have to simplify double sums like

$$\Sigma = \sum_{i,h\geq 0} 2^{(i+h)(2-N)}\cdot Q_\infty^2/Q_iQ_{i+h}.$$

This time we may use *Euler's product identity*

$$Q(t)=\sum_{n\geq 0}(-1)^n t^n/2^{\binom{n+1}{2}}Q_n =: \sum_{n\geq 0}a_{n+1}t^n \tag{5.15}$$

and find

$$\Sigma = \sum_{r\geq 0} a_{r+1}\, Q_{N+r-3}\cdot (1-2^{\,2-r-N})^{-1}.$$

In a similar manner it is possible to find an analytic function f(z) that fulfills (5.13):

$$f(z+1)=\sum_{r\geq 0} a_{r+1}\frac{Q}{Q(2^{3-z-r})}\cdot\left[2^z-\frac{2}{1-2^{1-z-r}}-\frac{2z}{1-2^{2-z-r}}+2\sum_{k\geq 2}\binom{z}{k}\frac{1}{2^{r+k-1}-1}\right] \tag{5.16}$$

As we have mentioned already the second main problem is to compute the mean of the periodic function $\delta_1^2(x)$, $\delta_1(x)$ from (5.9). We have for this mean (=zeroeth Fourier coefficient)

$$[\delta_1^2]_0 = \frac{1}{(\log 2)^2}\cdot\sum_{k\neq 0}\left|\Gamma\left(-1-\frac{2k\pi i}{\log 2}\right)\right|^2 \tag{5.17}$$

Now $|\Gamma(iy)|^2 = \pi/y\cdot\sinh(\pi y)$, so that $[\delta_1^2]_0$ may be expressed by series of the form $\sum_{k\geq 1}(k^m(e^{2\alpha k}-1))^{-1}$, m an odd natural $\geq 3$. Series of that type may be transformed via the following formula (that may be found in *Ramanujan's Notebooks*, compare *Berndt* [3]):

Let $\alpha$, $\beta > 0$ with $\alpha\cdot\beta=\pi^2$. Then

$$\begin{aligned}\alpha^{-N}\left(\frac{1}{2}\zeta(2N+1)+\sum_{k\geq 1}\frac{k^{-2N-1}}{e^{2\alpha k}-1}\right) &= (-\beta)^{-N}\left(\frac{1}{2}\zeta(2N+1)+\sum_{k\geq 1}\frac{k^{-2N-1}}{e^{2\beta k}-1}\right)\\ &\quad -2^{2N}\sum_{k=0}^{N+1}(-1)^N\frac{B_{2k}}{(2k)!}\frac{B_{2N+2-2k}}{(2N+2-2k)!}\alpha^{N+1-k}\beta^k,\end{aligned} \tag{5.18}$$

where $B_N$ is the N-th Bernoulli number.

Using (5.18) it can be proved that the terms of order $N^2$ in the variance of the path length cancel. Finally one gets

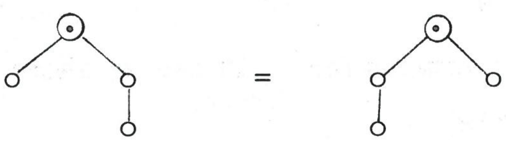
$$V_N \sim 0{,}26600\ldots\cdot N+\delta_3(\log_2 N)\cdot N+\mathcal{O}(N^{\varepsilon}). \tag{5.19}$$

A similar technique allows to analyze the average case behaviour of other relevant parameters of DST, as well as of Tries and Patricia Tries. (Compare [36] for further references.)

## 6. NONPLANAR TREES

### 6.1 Rooted Trees

In this section we consider the family $R$ of nonplanar rooted trees, i.e. rooted trees where the left-right order of the subtrees is irrelevant:



We have the constructive description

$$R = \{o\}\times M[R] \tag{6.1}$$

where $M[\cdot]$ is the multiset construction (compare equation (2.5)), so that the ogf fulfills

$$R(z) = z\cdot\exp(R(z)+\frac{R(z^2)}{2}+\frac{R(z^3)}{3}+\ldots), \tag{6.2}$$

a classical result by *Polya* (compare [23] for references and more detailed information on this section).

Let $\mathcal{R}_n$ be the family of n-node rooted trees. Then we have

$$\begin{aligned} \mathcal{R} &= \{o\} \times M[\mathcal{R}_1] \times M[\mathcal{R}_2] \times M[\mathcal{R}_3] \times \ldots \\ &= \{o\} \times (\underset{t\in\mathcal{R}_1}{\mathrm{X}} \{t\}^*) \times (\underset{t\in\mathcal{R}_2}{\mathrm{X}} \{t\}^*) \times (\underset{t\in\mathcal{R}_3}{\mathrm{X}} \{t\}^*) \times \ldots \end{aligned} \tag{6.3}$$

from which *Cayley's* result

$$R(z) = z\cdot\left(\frac{1}{1-z}\right)^{r_1}\cdot\left(\frac{1}{1-z^2}\right)^{r_2}\cdot\left(\frac{1}{1-z^3}\right)^{r_3}\ldots \tag{6.4}$$

is immediate. Starting from (6.2) we get the recursion

$$r_{n+1} = \frac{1}{n}\sum_{j=1}^{n}\Big(\sum_{d|j} d r_d\Big) r_{n-j+1}, \quad n\geq 1;\ r_1 = 1 \tag{6.5}$$

which allows to compute the first values of $r_n$:

| $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $r_n$ | 1 | 1 | 2 | 4 | 9 | 20 | 48 | 115 | 286 | 719 |

.

An asymptotic evaluation of $r_n$ may be gained along the following lines:

(i) $r_n \leq p_n$, the number of n-node planted plane trees, so that $R(z)$ has radius of convergence $q \geq 1/4$.

(ii) Let $f(z,y):=z\cdot\exp(y+\frac{R(z^2)}{2}+\frac{R(z^3)}{3}+\ldots)-y$.

Then $y=R(z)$ is the unique solution analytic around 0 of $f(z,y)=0$. The singularities on $|z|=q$ occur for $\frac{\partial f}{\partial y}(z,y)=0$:

$$\frac{\partial f}{\partial y} = z\cdot\exp(y+\frac{R(z^2)}{2}+\frac{R(z^3)}{3}+\ldots)-1 = f(z,y)+y-1, \tag{6.6}$$

so that the singularities $\eta$ with $|\eta|=q$ must fulfill

$$f(\eta,R(\eta))+R(\eta)-1 = 0\ . \tag{6.7}$$

(iii) Using relatively weak estimates for $r_n$ it can be shown that $R(\eta)$ exists and that $R(\eta) = \lim\limits_{\substack{z\to\eta \\ |z|<q}} R(z)$ (6.8)

(iv) From (6.7) and (6.8) we find

$$R(\eta) = 1 \quad . \tag{6.9}$$

From (6.9) together with (6.2) for $z=\eta$ one finds the numerical approximation

$$\eta = q = 0{,}338\ 3219\ldots, \qquad \text{resp.} \quad \eta^{-1} = 2{,}95576\ldots \tag{6.10}$$

(v) From (6.6) we get

$$\left.\frac{\partial^2 f}{\partial y^2}\right|_{\substack{z=\eta \\ y=R(\eta)=1}} = f(\eta,R(\eta))+R(\eta)=1 \neq 0.$$

Therefore it can be concluded that $R(z)$ allows an expansion

$$R(z)=R(\eta)-a_1(\eta-z)^{1/2}+a_2(\eta-z)+\ldots \ , \tag{6.11}$$

so that (compare (3.11))

$$r_n \sim \frac{a_1}{2\sqrt{\pi}}\eta^{-n+1/2} \cdot n^{-3/2}, \ n\to\infty \ . \tag{6.12}$$

In order to determine $a_1$ we use

$$R'(z)(1-R(z)) = \frac{1}{2}a_1^2 + c_1(\eta-z)^{1/2}+\ldots$$

so that

$$a_1^2/2 = \lim_{z\to\eta} R'(z)(1-R(z)). \tag{6.13}$$

Together with (6.2) this yields

$$\frac{a_1^2}{2} = \frac{1}{\eta} + \sum_{k\geqq 2} R'(\eta^k)\cdot \eta^{k-1},$$

whence

$$a_1 = 2{,}681127\ldots \ .$$

Altogether we have

$$r_n=0{,}4399237\ldots \cdot (2{,}95576\ldots)^n \cdot n^{-3/2}(1+O(\tfrac{1}{n})), \tag{6.14}$$

as has been proved by *Otter* [ 47].

## 6.2 Free Trees

Let F denote the family of *free*, i.e. unrooted, nonplanar, unlabelled, *trees*. We look for a possibility to express the ogf $F(z)$ in terms of $R(z)$, where $R(z)$ is the ogf of rooted trees (see Section 6.1; compare [23 ] for the following.)

For $t \in F$ let $\Gamma(t)$ be the automorphism group of $t$ and $v^*(t)$ the number of *"dissimilar nodes"*, i.e. of orbits of nodes of t under $\Gamma(t)$. Then

$$v_n = \sum_{t \in F_n} v^*(t) = r_n, \text{ resp. } V(z) = \sum v_n z^n = R(z). \qquad (6.15)$$

Let $e^*(t)$ denote the number of *"dissimilar edges"*, i.e. orbits of edges of t under $\Gamma(t)$. Then

$$e_n = \sum_{t \in F_n} e^*(t) = \text{the number of n-node trees "rooted" at an edge.}$$

Therefore

$$E(z) = \sum e_n z^n = \frac{1}{2} R(z)^2 + \frac{1}{2} R(z^2). \qquad (6.16)$$

We need a *relation between* $v^*(t)$ *and* $e^*(t)$:

Let $E_1, E_2, \ldots, E_{e^*(t)}$ be the orbits of edges of t under $\Gamma(t)$ and $v_i^*(t)$ the number of dissimilar endnodes of the edges in $E_i$. Then

$$v^*(t) - 1 = \sum_{i=1}^{e^*(t)} (v_i^*(t) - 1). \qquad (6.17)$$

The proof of (6.17) is by induction on $e^*(t)$:

For $e^*(t)=1$ we have $v^*(t)-1 = v_1^*(t)-1$.

For $e^*(t) \geqq 2$ we choose the orbit of an ending edge of the tree, w.l.o.g. $E_1$, remove $E_1$ from t (without the cutpoints!) and get a tree t' with $e^*(t)-1$ classes of edges and $v^*-(v_1^*-1)$ classes of nodes, so that the equality follows inductively.

Now $v_i^*(t)$ may only take the values 2 or 1 (where the latter instance corresponds to edges with two similar endpoints, i.e. *"symmetry lines"*).

Therefore we get from (6.17)

$$v^*(t) - 1 = e^*(t) - s^*(t), \qquad (6.18)$$

where $s^*(t)$ is the number of symmetry lines. (A tree has 0 or 1 symmetry lines, and $s^*(t)=1$ iff t is bicentered with the two central points in the same orbit.) From (6.18)

$$v_n = \sum_{t \in F_n} v^*(t) = \sum_{t \in F_n} 1 + e_n - s_n,$$

so that

$$V(z) = F(z) + E(z) - R(z^2). \qquad (6.19)$$

Together with (6.15) and (6.16) we finally have the desired formula

$$F(z) = R(z) - \frac{1}{2} R(z)^2 + \frac{1}{2} R(z^2). \qquad (6.20)$$

The first few values of $f_n$ are given below.

| n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $f_n$ | 1 | 1 | 1 | 2 | 3 | 6 | 11 | 23 | 47 | 106 |

In order to get an asymptotic estimate we may use the analysis of the previous section:

$F(z)$ has again the dominating singularity $\eta$. From (6.11)

$$\frac{1}{2} R(z)^2 = \frac{1}{2} - a_1(\eta-z)^{1/2}+\ldots$$

while $R(z^2)$ is analytic in $z=\eta$. Therefore in $F(z)$ the $(\eta-z)^{1/2}$-term cancels and $(\eta-z)^{3/2}$ is the dominating term. A calculation of the constants finally yields

$$f_n \sim 0{,}5349\ 485 \cdot \eta^{-n} \cdot n^{-5/2}, \quad n\to\infty \tag{6.21}$$

(*Otter*, [47].)

Comparing with (6.15) we find

$$\lim_{n\to\infty} \frac{r_n}{nf_n} = \frac{1}{1+\sum\limits_{k\geq 2} R'(\eta^k)\cdot\eta^k} = 0{,}822366\ldots \quad . \tag{6.22}$$

## 7. LABELLED TREES

### 7.1 The Combinatorics of the Exponential Generating Function

For the study of labelled trees it is convenient to give some general remarks on the *operator method* for exponential generating functions (egf) of labelled objects: Again we follow *Flajolet's* approach [17].

Let $A$ be a class of *labelled objects* with a weight function, where the objects $t\in A_n$ (i.e. of size n) are labelled with $\{1,2,\ldots,n\}$. If $a_n=|A_n|$ we have the egf

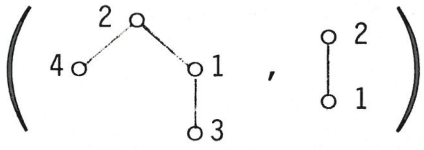
$$\hat{A}(z) = \sum_{n\geq 0} a_n z^n/n! \tag{7.1}$$

1) The *disjoint union* $A \dot{\cup} B$ of families of labelled objects has egf $\hat{A}(z)+\hat{B}(z)$.

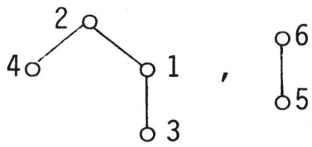
2) The combinatorial construction corresponding to the *Cauchy product* $\hat{A}(z)\cdot\hat{B}(z)$ is $A*B$ defined as follows: The elements of $A*B$ are all ordered pairs $(t_1,t_2)$, $t_1\in A$, $t_2\in B$, relabelled with the numbers

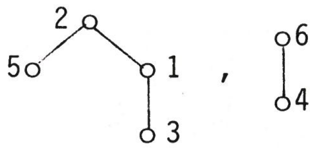
$\{1,2,\ldots,|t_1|_A+|t_2|_B\}$ in the following way. Take all bipartitions of $\{1,2,\ldots,|t_1|_A+|t_2|_B\}$ into a set $\{\alpha_1,\alpha_2,\ldots,\alpha_{|t_1|_A}\}$ of size $|t_1|_A$ and a set $\{\beta_1,\beta_2,\ldots,\beta_{|t_2|_B}\}$ of size $|t_2|_B$ and replace in $t_1$ the label i by $\alpha_i$ $(1\leq i\leq |t_1|_A)$ and in $t_2$ the label i by $\beta_i$ $(1\leq i\leq |t_1|_B)$. The set of all accordingly relabelled pairs $(t_1,t_2)$ is $A*B$.

<u>Example:</u> A labelled tree, is a (free) tree t where the nodes are labelled by $1,2,\dots,|t|$. The correct relabellings of a pair

$(t_1,t_2) =$  are formed as follows. We have

$|t_1|+|t_2| = 6$, so that we have to consider all $\binom{6}{4} = 15$ bipartitions of $\{1,2,\dots,6\}$ of type (4,2). The first 2 bipartitions yield the following relabellings:

$\{1,2,3,4\}\ \{5,6\} \rightarrow$ 

$\{1,2,3,5\}\ \{4,5\} \rightarrow$ 

3) If we set $A^{<k>} = A * A * \dots * A$ (k-times), then the egf is $\hat{A}(z)^k$.

4) The *"partitional complex of* A" is $A^{<*>} = \dot{\bigcup\limits_{k\geq 0}} A^{<k>}$ with $A^{<0>} = \{\varepsilon\}$ $(a_0=0)$ and has the egf $\dfrac{1}{1-\hat{A}(z)}$.

5) $A^{[k]} = \{\{t_1,\dots,t_k\} \mid (t_1,t_2,\dots,t_k) \in A^{<k>}\}$, i.e. the k-element multisets of objects of A with correct relabelling, have egf $\hat{A}(z)^k/k!$ .

6) $A^{[*]} = \dot{\bigcup\limits_{k\geq 0}} A^{[k]}$ $(a_0=0)$, the *"Abelian partitional complex"* has egf $\exp(\hat{A}(z))$.

## <u>7.2 Labelled Trees</u>

Let $L$ denote the family of *labelled trees* (i.e. general trees t with nodes labelled $1,2,\dots,|t|$) and $L_R$ the family of *rooted labelled trees*.

Then we have (with the symbols of Section 7.1)

$$L_R = \{o\} * L^{[*]} \ , \tag{7.2}$$

since each rooted labelled tree may be cut down at the edges following the root yielding a single node (the former root) and an abelian complex of rooted labelled trees (the new roots formed by the nodes adjacent to the former root). Therefore the egf fulfills

$$\hat{L}_R(z) = z \cdot \exp(\hat{L}_R(z)) \ . \tag{7.3}$$

By Lagrange inversion we find

$$l_{n,R}/n! = [z^n]\,\hat{L}_R(z) = \frac{1}{n}[z^{n-1}]\,e^{nz} = n^{n-1}/n!\ ,$$

so that

$$l_{n,R} = n^{n-1} \qquad (\textit{rooted, labelled})$$

resp. (7.4)

$$l_n = n^{n-2} \qquad (\textit{labelled}).$$

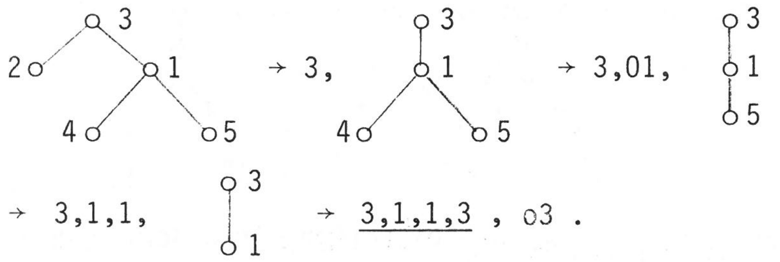
Formulas (7.4) are usually attributed to *Cayley* [6], but were in fact already known ealier (an equivalent result was proved 1860 by *Borchardt*, and the result appeared without proof already in 1857 in a paper by *Sylvester*, compare *Moon* [ 43] for references). In [42] Moon has presented several proofs for (7.4). We want to present here two combinatorial proofs:

1) The *Prüfer code* [ 50]

Let t be a tree in $L_R$ and f(v) denote the label of the "father" of node labelled v in t.

We start by taking the endnode $v_1$ with smallest number, note $f(v_1)$ and remove $v_1$ together with the edge incident to $v_1$. Then we recursively proceed with the remaining tree.

Example:




The result is a sequence of length n-1 with elements in {1,2,...,n}, which is called the *Prüfer code* of the tree. It is easily seen that this map

$$L_{n,R} \to \{1,2,\ldots,n\}^{n-1}$$

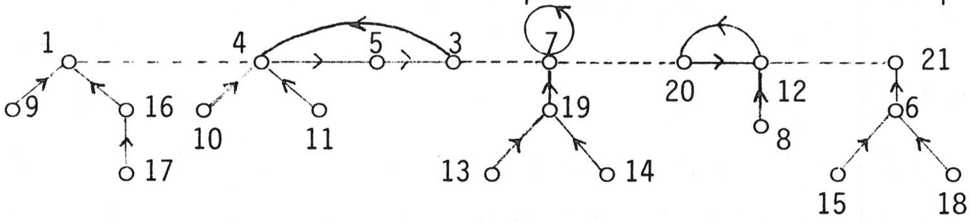
is bijective (the reader may immediately reconstruct the tree from the code). Therefore we find again (7.4).

2) There are other *bijections that preserve* certain *weights* on the trees, and are therefore of a particular interest. We present here the construction of *Eğecioğlu* and *Remmel* [16 ] which gives a bijection between $\mathcal{F}_{n+1}$, the set of all functions $\{2,\ldots,n\} \to \{1,2,\ldots,n+1\}$ and $L_{n+1,n+1}$ the set of labelled trees with n+1 nodes, rooted at n+1:

Let us, for example, consider the following function f in $\mathcal{F}_{21}$:

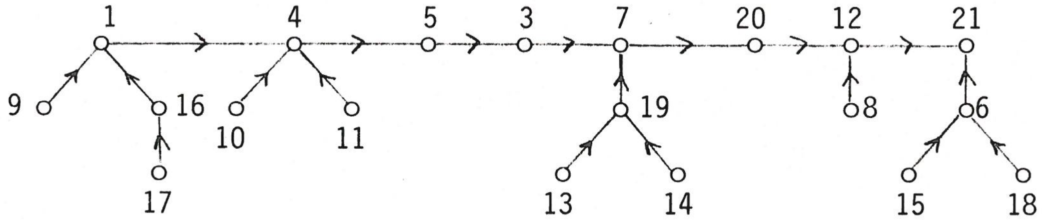
| i | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| f(i) | 5 | 4 | 5 | 3 | 21 | 7 | 12 | 1 | 4 | 4 | 20 | 19 | 19 | 6 | 1 | 16 | 6 | 7 | 12 |

We build up the *directed graph* $G_f$ of f, where $<i,j> \in E(G_f)$ iff $f(i)=j$:




The components are i) 2 trees rooted at 1 resp. n+1 drawn at the extreme left and right, and ii) directed cycles of length $\geq 1$, where for each vertex v a tree rooted at v may be attached. These cycles shall be drawn as directed paths on the line connecting 1 and n+1, with one additional edge "backwards" on the top, the trees below, the smallest elements of cycles at the right, the cycles ordered from left to right by increasing smallest elements.

In the next step we delete the "backward" edges $\overset{\curvearrowleft}{l_i \quad r_i}$ from the top and add the new edges $<1,l_1>,\dots,<r_i,l_{i+1}>,\dots,<r_k,n+1>$. The resulting digraph is a tree in $L_{n+1,n+1}$:




The whole mapping $\Theta_{n+1}: \mathcal{F}_{n+1} \to L_{n+1,n+1}$ is a bijection: In order to reconstruct the function from a given tree in $L_{n+1,n+1}$, consider $r_1$, the smallest element on the path between 1 and n+1, $r_2$, the smallest element on the path between $r_1$ and n+1, and so on. Then $r_i$ are the rightmost elements of the cycles in the directed graph of f, which can easily be reconstructed from this information.

The bijection $\Theta_{n+1}$ is *weight preserving* in the following sense, and therefore allows to prove a *q-analogue of Cayley's formula* (7.4):

For $t \in L_{n+1,n+1}$ we associate a weight w to each directed edge (all edges directed towards n+1), where $w(<i,j>) = q^i u^j$ if $i > j$ resp. $yp^i s^j$ if $i < j$.

Furthermore $w(t)=\prod_{e\in E(t)} w(e)$. Then the following identity holds

$$\sum_{t\in L_{n+1,n+1}} w(t) = yps^{n+1}\cdot\prod_{i=2}^{n}[xq^i(u+u^2+\ldots+u^{i-1})+yp^i(s^i+\ldots+s^{n+1})]. \tag{7.5}$$

In order to prove (7.5) we define a weight $\sigma$ on $\mathcal{F}_{n+1}$ by $\sigma(f)=\prod_{i=2}^{n}\sigma(f,i)$ where $\sigma(f,i)=xq^iu^j$ if $f(i)=j$ and $i>j$, resp. $yp^is^j$ if $f(i)=j$ and $i\leq j$. Then

$$\sum_{f\in\mathcal{F}_{n+1}}\sigma(f)=\prod_{i=2}^{n}[xq^iu+xq^iu^2+\ldots+xq^iu^{i-1}+yp^is^i+\ldots+yp^is^{n+1}].$$

Therefore it is sufficient to prove

$$w(\Theta_{n+1}(f)) = yps^{n+1}\sigma(f) \text{ for all } f\in\mathcal{F}_{n+1}. \tag{7.6}$$

We note that $w(<i,j>)=\sigma(f,i)$ if $f(i)=j$. The change of weights under $\Theta_{n+1}$ comes from the cancellation of the "backward" edges and the addition of the edges $<1,l_1>,\ldots,<r_i,l_i>,\ldots,<r_k,n+1>$. Since $r_i$ is the smallest element in the cycle we have $l_i=f(r_i)\geq r_i$, so that $\sigma(f,r_i)=yp^{r_i}s^{l_i}$ and

$$yps^{n+1}\sigma(f) = yps^{n+1}yp^{s_1}s^{l_1}\ldots yp^{r_k}s^{l_k}\cdot\prod\sigma(f,i),$$

where the product is over $i\notin\{r_1,\ldots,r_k\}$. The latter product equals $\prod w(<i,j>)$ over all "non-backward" edges $<i,j>$. The product over all "backward" edges is $yps^{l_1}yp^{r_1}s^{l_2}..yp^{r_k}s^{n+1}$, since $r_1<r_2<\ldots<r_k$, so that $r_i<l_{i+1}$. Altogether we get (7.6).

Some *consequences of* (7.5):

1) $x=y=1$, $p=s=u=q$. Then $w(<i,j>)=q^iq^j$, so that node $i$ contributes $q^{id(i)}$ to $w(t)$, where $d(i)$ is the (total) degree of $i$. Therefore $w(t)=q^{\delta(t)}$, with $\delta(t)=\sum_i id(i)$, and we get

$$\sum_{t\in L_{n+1}} q^{\delta(t)} = q^{(n^2+5n)/2}\cdot([n+1]_q)^{n-1}, \tag{7.7}$$

where $[n+1]_q=\dfrac{q^{n+1}-1}{q-1}$. This is a *q-analogue* of $|L_{n+1}|=(n+1)^{n-1}$.

2) $y=p=q=s=t=1$: We count by $x$ edges $<i,j>$ with $i>j$, i.e. the "falls" in $t\in L_{n+1,n+1}$ (edges directed towards $n+1$).

Thus $\sum_t x^{\#\text{ falls in } t} = (x+n)(2x+n-1)+\ldots+((n-1)x+2)$.

From this we may conclude, e.g., that the *average number of "falls" in a tree in* $L_{n+1,n+1}$ is $n(n-1)/2(n+1)$.

## References:


[1] G.ANDREWS, "The Theory of Partitions", Addison Wesley, Reading (1976).

[2] E.BENDER, Asymptotic methods in enumeration, SIAM Rev.16(1974),485-515.

[3] B.C.BERNDT, "Ramanujan's Notebooks, Part.II", Springer, New York(1989).

[4] N.G.DE BRUIJN, "Asymptotic Methods in Analysis", Dover, New York(1981).

[5] N.G.DE BRUIJN, D.E.KNUTH, S.O.RICE, The average height of planted plane trees, in "Graph Theory and Computing", R.-C.Read ed., Academic Press, New York(1972),15-22.

[6] A.CAYLEY, A theorem on trees, Quart.J.Math.23(1889), 376-378.

[7] E.G.COFFMAN Jr., J.EVE, File structures using hashing functions, Comm. ACM 13(1970), 427-436.

[8] L.COMTET, "Advanced Combinatorics", D.Reidel, Dordrecht(1974).

[9] G.DARBOUX, Memoire sur l'approximation des fonctions de trés grands nombres ..., J.Math.pures appl.4(1878), 5-56 and 377-416.

[10] N.DERSHOWITZ, S. ZAKS, Enumerations of ordered trees, Discrete Math.31 (1980), 9-28.

[11] N.DERSHOWITZ, S. ZAKS, Ordered trees and non-crossing partitions, Discrete Math.62(1986), 215-218.

[12] N.DERSHOWITZ, S. ZAKS, The Cycle Lemma and some applications, Europ.J.Comb. 11(1990), 35-40.

[13] G.DOETSCH,"Handbuch der Laplace-Transformation", Birkhäuser, Basel(1955).

[14] A.DVORETZKY, TH.MOTZKIN, A problem of arrangements, Duke Math.J.14(1947), 305-313.

[15] P.H.EDELMAN, Chain enumeration and non-crossing partitions, Discrete Math.31 (1980), 171-180.

[16] Ö.EĞECIOĞLU, J.REMMEL, Bijections for Cayley trees, spanning trees and their q-analogues, J.Comb.Th.A 42(1986), 15-30

[17] P.FLAJOLET, Mathematical methods in the analysis of algorithms and data structures, in: "Trends in Theoretical Computer Science" (E.Börger ed.), Computer Science Press, Rockville, Maryland(1988), 225-304.

[18] P.FLAJOLET, A.ODLYZKO, The average height of binary trees and other simple trees, J.Comput.System Sci.25(1982), 171-213.

[19] P.FLAJOLET, A.ODLYZKO, Singularity analysis of generating functions, SIAM J. Disc.Math.3(1990), 216-240.

[20] P.FLAJOLET, H.PRODINGER, Level number sequences for trees, Discrete Math.65 (1987), 149-156.

[21] P.FLAJOLET, R.SEDGEWICK, Digital search trees revisited, SIAM J.Comput.15 (1986), 748-767.

[22] I.P.GOULDEN, D.JACKSON,"Combinatorial Enumeration", John Wiley, New York(1983).

[23] F.HARARY, E.M.PALMER, "Graphical Enumeration", Academic Press, New York(1973).

[24] P.HENRICI, "Applied and Computational Complex Analysis, vol.I", J.Wiley, New York(1988).

[25] A.JOYAL, Une théorie combinatoire des séries formelles, Adv.in Math.42(1981), 1-82.

[26] R.KEMP, The average height of r-tuply rooted planted plane trees, Computing 25(1980), 209-232.

[27 ] R.KEMP, The average height of planted plane trees with m leaves, J.Comb. Th.B 34(1983), 191-208.

[28 ] R.KEMP, "Fundamentals of the Average Case Analysis of Particular Algorithms", J.Wiley, Chichester(1984).

[29 ] R.KEMP, On the number of deepest nodes in ordered trees, Discrete Math.81 (1990), 247-258.

[30 ] P.KIRSCHENHOFER, On the height of leaves in binary trees, J.Comb.Syst.Sci.8 (1983), 44-60.

[31 ] P.KIRSCHENHOFER, Some new results on the average height of binary trees, Ars Combinatoria 16A(1983), 255-260.

[32 ] P.KIRSCHENHOFER, A tree enumeration problem involving the asymptotics of the "diagonals" of a power series, Annals of Discrete Math.33(1987),157-170.

[33 ] P.KIRSCHENHOFER, Asymptotische Untersuchungen zur durchschnittlichen Gestalt gewisser Graphenklassen II, in "Zahlentheoretische Analysis II",E.Hlawka ed., Springer Lecture Notes in Math.1262(1987), 86-92.

[34 ] P.KIRSCHENHOFER, H.PRODINGER, On the recursion depth of special tree traversal algorithms, Information and Computation 74(1986), 15-32.

[35 ] P.KIRSCHENHOFER, H.PRODINGER, On some applications of formulae of Ramanujan in the analysis of algorithms, Mathematika, to appear.

[36 ] P.KIRSCHENHOFER, H.PRODINGER, W.SZPANKOWSKI, Digital seach trees again revisited: the internal path length perspective, preprint,TU Vienna,1990.

[37 ] D.E.KNUTH, "The Art of Computer Programming, vol.I", Addison Wesley, Reading (1973).

[38 ] D.E.KNUTH, "The Art of Computer Programming, vol.III", Addison Wesley, Reading(1973).

[39 ] A.G.KONHEIM, D.J.NEWMAN, A note on growing binary trees, Discrete Math.4 (1973), 57-63.

[40 ] G.KREWERAS, Sur les partitions non croisées d'une cycle, Discrete Math.1 (1972), 333-350.

[41 ] A.MEIR, J.MOON, On the altitude of nodes in random trees, Can.J.Math.30 (1978), 997-1015.

[42 ] J.MOON, Various proofs of Cayley's formula for counting trees, in:"A Seminar on Graph Theory", F.Harary ed.,Holt,New York(1967),70-78.

[43 ] J.MOON, "Counting Labelled Trees", Canadian Math.Monographs 1(1970).

[44 ] J.MOON, On level numbers of t-ary trees, SIAM J.Alg.Disc.Math.4(1983),8-13.

[45 ] T.V.NARAYANA, A partial order and its application to probability, Sankhya 21 (1959), 91-98.

[46 ] N.E.NÖRLUND, "Vorlesungen über Differenzenrechnung", Chelsea,New York(1954).

[47 ] R.OTTER, The number of trees, Ann.of Math.49(1948), 583-599.

[48 ] H.PRODINGER, A correspondence between ordered trees and noncrossing partitions, Discrete Math.46(1983), 205-206.

[49 ] H.PRODINGER, The average height of the d-th highest leaf of a planted plane tree, Networks 16(1986), 67-75.

[50 ] H.PRÜFER, Neuer Beweis eines Satzes über Permutationen, Arch.Math.Phys.Sci. 27(1918), 742-744.

[51 ] A.RENYI, G.SZEKERES, On the height of trees,Austral.J.Math.7(1967),497-507.

[52 ] N.J.A.SLOANE, "A Handbook of Integer Sequences",Academic Press,New York(1973).

[53 ] V.STREHL, Two short proofs of Kemp's identity for rooted plane trees, Europ.J. Comb.5(1984), 373-376.